

THIS WEEK

EDITORIALS

GENOME Bizarre creature the naked mole rat has its sequence published **p.156**

WORLD VIEW Arctic regulation cannot wait until the sea ice goes **p.157**

DEATH ROW Cancer cull won't save Tasmanian devils **p.158**



The case of Dr Hicheur

The continued imprisonment of a French–Algerian physicist highlights the need for scientists to defend the human rights of all colleagues.

Is Adlène Hicheur a dangerous terrorist? The French government certainly suspects so. When he was arrested in 2009, authorities said that Hicheur, then a postdoc in high-energy physics at the Swiss Federal Institute of Technology in Lausanne, was plotting terrorist attacks in France. Since then, they have held him in custody, without trial, at Fresnes Prison near Paris, while they gather further evidence.

Hicheur's family and close colleagues say the idea that Hicheur planned specific terrorist attacks is false and fantastical. They say that the French–Algerian physicist was involved in spirited online political debates about many topics, including terrorism. They argue that his arrest was a political message sent by the government to France's Muslim community, warning that everyone, no matter how educated, is under scrutiny.

Hicheur's work at CERN, Europe's premier high-energy physics laboratory near Geneva in Switzerland, certainly made for some good headlines — “Nuclear terror suspect is top physicist”, for example. And, at this stage, *Nature* is in no position to judge his innocence or guilt, which is a question for the French judicial system.

Nonetheless, Hicheur's case deserves attention because to have held him in custody for so long, although legal under France's tough anti-terror laws, seems a clear abuse of human rights.

For the two years since his arrest, his online exchanges, which took place in web forums and chat rooms, have remained undisclosed by French authorities. Hicheur and his lawyers are eager to share what he said, but they are forbidden by law from doing so. He remains locked away and largely forgotten. This week a judge, who under the French system is independent of the police and prosecution, was scheduled to close a ‘preliminary investigation’ and hand the case to prosecutors, who will determine whether a trial should go ahead. Under French law, Hicheur could be held for a further 12 months without his case being heard, his defence team says.

Persecution of scientists, and physicists in particular, is nothing new. During the cold war, researchers on both sides of the iron curtain suffered for their political views. In the United States, Robert Oppenheimer's career was ruined by rumours of communist sympathies. And Soviet scientist and dissident Andrei Sakharov spent much of the 1980s in internal exile for his outspoken views on human rights and arms control.

Hicheur is part of a new generation. Even as his incarceration stretches into its third year, Omid Kokabee, a physicist trained in the United States and Spain, awaits trial in Iran on charges of “communicating with a hostile government”. Also in Iran, several physicists have been mysteriously assassinated in recent years by unknown agents presumably trying to slow the nation's nuclear programme.

These new cases are related to terrorism, not communism, but the similarities are abundant. As in the 1950s and 60s, there is the fear of an internal threat — a ‘sleeping cell’ that could activate at any time. The spread of technology to hostile governments is also a worry. Where once there was Klaus Fuchs, the German-born scientist turned Soviet

spy, now there is Abdul Qadeer Khan, the Pakistani atomic-weapons researcher who ran a nuclear-trafficking ring.

However, there is one important difference. During the cold war, many scientists spoke out in defence of their imprisoned peers; today, the scientific community remains largely silent. Hicheur has had support from his closest colleagues, but CERN, where he once worked, has

done its best to distance itself from him. Most scientific societies in France and Europe seem uninterested in the case. This indifference is part of a larger trend — in the United States, scientific organizations are eschewing human-rights campaigns for individual scientists in favour of broader (and blander) programmes of ‘scientific cooperation’ (see *Nature* **475**, 431–432; 2011).

There are two reasons why this might be. The scientists persecuted during the cold-war era were senior researchers, whereas today's victims tend to be early-career postdocs and graduate students. Moreover, scientists on both sides of that earlier conflict were bound by national origins, ethnicity and religion — something that seems less true today. Western scientists should also ask themselves whether they are less willing to speak out on Hicheur's behalf because he is a practising Muslim.

There has never been a more important time to rally behind scientists such as Hicheur. This year has seen upheaval across the Arab world, as citizens throw off their oppressive regimes. As educated, free-thinking members of their societies, scientists can play a part in these political upheavals. These researchers can help to open their societies, and serve as a vital bridge between the Muslim world and the West.

Hicheur may have been just such a bridge. In his online debates, he says, he discouraged acts of terrorism against innocent civilians. Whether or not that is true, he should be released until his trial. ■

An eye for success

Steve Jobs and Apple revolutionized the way scientists render their work.

Much of the praise heaped on Steve Jobs, the chief executive and co-founder of computer firm Apple, came from those who were introduced to the man and his company through the gadget wizardry of the iPhone and iPad. And it is probably through the eyes of this ‘iGeneration’ that the legacy of Jobs, who died last week, will be sealed. Yet Jobs and Apple did more than just revolutionize the way that people fill their spare time.

The use of computers in science now may bring images of

data-crunching parallel processing and Unix-inspired open-source collaborations, but there was a time when the cutting edge — at least for molecular scientists — was being able to draw a benzene ring digitally on screen. Jobs and Apple gave scientists the power to do that, and more besides. Scientists responded with decades of loyalty and, on news of Jobs's death, with tributes of their own.

Apple lost the battle for dominance of corporate computing to the PC long ago, yet there is a good reason for the devotion to Apple computers in industries such as design and publishing (even if it might seem irrational to outsiders). The introduction in 1984 of the Apple Macintosh, or Mac, complete with mouse, brought easy graphic imaging to the masses. In doing so, it slashed the time taken for chemists and molecular biologists, among others, to represent their work and deliver their papers. Windows as we know it was years away, and Apple had a head start in these fields that, for many scientists, it never relinquished.

The mid-1980s was a time ruled by technology giants such as the Grafacon 1010, a table-top-sized digitizing tablet that was operated by a foot pedal. For the scientists fortunate enough to have access, the machine converted hand drawings into glowing lines on a screen. Most researchers got by with sheets of Letraset letter and symbol transfers bought from the local stationer and painstakingly rubbed onto written pages at the correct positions. Foreshadowing the modern use of punctuation to portray emotions (emojis), chemists were not averse to inserting sheets of paper into their typewriters at oblique angles and using the equals sign and dashes to represent chemical bonds.

Henry Rzepa, a computational chemist at Imperial College London, says that the Mac, in tandem with associated software such as ChemDraw and early laser printers, “gave us back our science”. Before its introduction, he says, it was not unusual for a university scientist to spend an entire week transferring all the letters and symbols onto the pages of a research paper, which journal publishers were demanding in ‘camera-ready’ form. Some academic departments employed a staff member to do nothing else. The Mac

made it a five-minute job and, Rzepa says, the effect was liberating. (Certainly, in a sign of what future Apple technology would do for record shops, the unfortunate staff member who was skilled with the transfers often found themselves liberated from a job.)

Susan Forsburg, professor of molecular and computational biology at the University of Southern California in Los Angeles, was an early user of the Mac. “It’s hard to explain to my students today what a change it was to go from laboriously hand-drawing figures to making clean, crisp, computer-generated line drawings and digital images,” she says. “It was revolutionary.” Molecular biologists have been attracted to Apple technology ever since, she adds, and at least half of the attendees at scientific meetings are likely to have one on their desk. “These days, a lot of graphics software runs on Windows but I think Macs still have the edge, in part because they focus so much on ease of use. But I would think that, wouldn’t I?”

Advocates of the PC are no doubt gnashing their teeth at this point, and rose-tinted nostalgia has probably helped to delete memories of some of the early Mac’s shortcomings. A comprehensive review of “software for two-dimensional chemical structure editing” published in the late 1980s (A. C. Norris and A. L. Oakley *Comput. Chem.* **12**, 245–251; 1988), for instance, rather stiffly points out that the Macintosh version of ChemDraw “is more difficult for a chemist to use than is at first apparent”. Among the drawbacks: “Rings have to be sized and oriented each time they are drawn, making drawing of fused rings tedious.” Still, this did not stop the potential of the new trend being recognized by the authors, who concluded: “Electronic publishing is likely to have a considerable impact on the way in which authors, research and commercial concerns, and publishers produce and distribute information.” Quite — and Jobs and Apple more than played their part. Thank you, Steve. ■

More than teeth

The bizarre-looking naked mole rat is a worthy member of the genome club.

In an unflattering light, a naked mole rat looks like a wrinkly sausage with oversized teeth, legs and a tail. And given that it spends all of its extraordinarily long life short of air in dark and overcrowded underground tunnels, where it frequently eats its own excrement, an unflattering light is probably the best that a naked mole rat can hope for. Still, the best science, like love and justice, is blind, so this week the naked mole rat (*Heterocephalus glaber*; also known as the sand puppy or desert mole rat) joins the illustrious list of animals judged to be of sufficient significance for an analysis of their genome sequence to be published.

And what an animal it is. Unfortunately, the research paper that describes its genetic insides, published online by *Nature* this week (E. B. Kim *et al.* *Nature* doi:10.1038/nature10533; 2011), finds no room to feature a clear image of its extraordinary outsides. So those readers unfamiliar with this bizarre burrowing rodent native to parts of east Africa are highly recommended to look up its image on the Internet. See what we mean about the teeth?

And there’s more: the mouse-sized creature is one of only two mammals known to live in ant- and termite-style eusocial colonies (the other being the Damaraland mole rat). A naked mole rat queen suppresses the sexual maturity of her subordinates, and on a royal

death, the female who wins the fight to take the queen’s place must stretch herself to pup-bearing size. At length, she is joined by a select few breeding males, while the other members of the colony — sterile workers — dedicate their days to the search for food and to digging and defending tunnels. Should it need to, a naked mole rat soldier can shuffle backwards as fast as its little feet normally carry it forwards.

The creature is certainly an interesting curiosity, but in these cash-strapped times, why invest in the secrets of its genome?

The naked mole rat is one of Mother Nature’s great survivors. The busy underground lairs in which the animals live almost always run low on oxygen and high on carbon dioxide. Steady subterranean temperatures have sapped the creatures’ ability to regulate their body temperature. Yet what they sacrifice in quality of life they more than make up for in extraordinary quantity. Comfortably the longest-living rodent, naked mole rats can live for more than 30 years. They seem impervious to cancer and do not feel some types of pain.

All of which means that the frankly ugly naked mole rat could prove a sight for sore eyes in the biomedical community. The information published on its genome and transcriptome has already revealed patterns of gene expression different from those in humans, mice and rats, and this may underlie its longevity. With further study, mechanisms of ageing, genetic regulation of lifespan, adaption to extreme environments, low-oxygen tolerance, cancer resistance, sexual development and hormonal regulation are up for grabs.

As those reporting the genome sequence say, it provides a “rich resource that can be mined in numerous ways to uncover the molecular bases for the extraordinary traits of this most unusual mammal”. And did we mention the teeth? ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Marine protection in the Arctic cannot wait

Global economics, not declining sea ice, is driving ships to the Arctic Ocean. Only international regulation will protect the region, says Lawson Brigham.

Most people know that profound change is happening in the Arctic Ocean. Most people would say that this is because the sea ice there is in retreat. But most people would be wrong. Changes in ice cover are only part of a story that is, in fact, driven largely by economics and geopolitics. Despite the headlines, policy-makers, planners and regulators need to look beyond the disappearing ice and understand the economic drivers to grasp the urgent need for maritime regulations to address the booming development of the region.

It is true, of course, that Arctic sea ice is in decline. In the summer of 1994, I was part of a historic scientific crossing of the Arctic Ocean, from the Bering Strait to the North Pole and out the other side through the Fram Strait. Sailing in company, the icebreakers *Polar Sea* and *Louis S. St-Laurent* encountered nearly 2,000 nautical miles of continuous sea ice. I remember, because *Polar Sea*, of which I was captain, observed, sampled and broke through much of it. A similar expedition this summer would probably have seen ice for only half that distance — and much of it would have been thinner.

It is also true that retreating ice provides greater access throughout the Arctic Ocean. This has led to speculation about quicker trade routes across the entire maritime Arctic. But what tends to be overlooked in these discussions about trans-Arctic navigation is that traffic is already increasing in the north, driven by marine tourism and the development of natural resources.

This reality must be understood because it makes the issue of how to improve protection of Arctic people and the environment a question not for the future, but for now.

The Arctic is connecting to the global economy, so development there is being driven by global commodity prices. Already we see summertime shipping from the world's largest zinc mine (Red Dog), off the northwest coast of Alaska; in northwest Russia icebreaking container ships navigate year-round to near the world's largest nickel mine in Norilsk; icebreaking shuttle tankers ply offshore facilities in the Pechora Sea; and summer oil drilling off western Greenland brings drill ships and a support fleet. The past two summers have seen large tankers and bulk carriers sailing across the Russian maritime Arctic to investigate how to carry natural resources from Russia and northern Norway to China. Also, a fleet of icebreaking ships is planned to take high-grade iron from a mine on Canada's Baffin Island to steel producers in Europe. And more and more tourists are visiting the region aboard cruise ships, especially along Greenland's west coast.



AFTER THE ICE

Science at the new Arctic frontier
nature.com/arcticfrontier

The rise in ship traffic has the potential to increase pollution in the region and heightens the risk of spills. We therefore need to develop effective international rules and regulations to enhance marine safety and protect the environment.

In December 2009, the International Maritime Organization (IMO), the United Nations agency responsible for the safety and security of shipping and the prevention of pollution by ships, developed voluntary guidelines for ships operating in polar waters that recognize the unique challenges of the region. Temperatures there are low and weather conditions poor. There is also a lack of adequate charts and other navigation and communication services, and the waters are remote from search and rescue, salvage and other maritime infrastructure found at lower latitudes. Sea ice also presents specific stresses on ship hulls and systems.

The IMO must make such standards mandatory. Indeed, that was the conclusion of Arctic Council's *Arctic Marine Shipping Assessment Report 2009*. The council — an inter-governmental forum of the eight states with territory in the Arctic — called for structural standards for ships sailing to the Arctic, the carriage of adequate marine safety equipment and enhanced training for ice navigators. Arctic ministers accepted these recommendations and the Arctic states, led by Norway, are working with the IMO to make them mandatory. A binding set of standards is due to be introduced by 2013. Even then, however, the lack of infrastructure will continue to undermine safe and efficient Arctic shipping. We need more public

and commercial investments, and new public-private ventures, to rapidly develop this crucial safety net for the region.

Loss of sea ice potentially allows longer seasons for scientific research, commercial operations and even adventurers. And the possibility of a brief, ice-free period in summer throughout the Arctic Ocean, which some researchers expect to happen within two decades, will be a remarkable physical change. Most climate models show that despite the warming, the Arctic Ocean will be fully or partially ice-covered for a good part of the year. Thus, all ships are likely to encounter sea ice.

If the maritime Arctic is to be developed safely, then greater attention to detail is needed. International cooperation, better science and knowledge of the local environment and investment in key infrastructure will be essential. ■

Lawson Brigham is distinguished professor of geography and Arctic policy at the University of Alaska Fairbanks, and former chair of the Arctic Council's Arctic Marine Shipping Assessment.
e-mail: lwbrigham@alaska.edu

SEVEN DAYS

The news in brief

POLICY

Tar-sands disputes

Europe is moving to block the importation of oil from tar sands in Canada by revising its fuel directive to classify the fuel source as highly polluting in terms of carbon dioxide emissions. Because the directive requires fuel importers to reduce emissions, the classification would effectively shut down a European market for the sources, and Canada's natural-resources minister Joe Oliver said that his country would complain to the World Trade Organization if the directive — passed last week by the European Commission — is approved by member states. Canada's main market for tar-sands oil will be the United States, where the government last week faced contentious final public hearings before it decides whether to approve a 2,700-kilometre pipeline from Alberta in Canada to Texas.

US geoengineering

A panel of scientists, engineers, environmentalists and policy experts has urged the US government to coordinate a



S. GARDNER/REUTERS

Cleaning up the Gulf Coast

Last year's oil spill was just the latest disaster to hit the Gulf of Mexico, which has suffered years of environmental damage. Now a federal task force has laid out a long-term strategy to restore ecosystems in the region. The 5 October preliminary report, released a year after it was commissioned by US President Barack Obama, focuses on goals such as restoring wetlands and water quality; replenishing marine

resources; and enhancing the resilience of Gulf Coast communities. The final version will be published in December. According to bills that have been proposed but not yet passed in both arms of Congress, around 80% of the US\$5-billion to \$20-billion fines collected after the oil spill could go directly towards the restoration effort, with perhaps 5% of that for new science programmes.

multi-agency programme of research into the potential of geoengineering, or 'climate remediation' technologies, as the panel termed such techniques. The task force, organized by the Bipartisan Policy Center in Washington DC, published its report on 4 October, just five days after the European Parliament passed a resolution expressing "opposition to proposals for large scale geo-engineering". See go.nature.com/7fxfb for more.

War on Guinea worm

The Carter Center in Atlanta, Georgia, has announced a new push to wipe out Guinea worm disease (dracunculiasis) by 2015. At an appeal in London on 5 October, the centre, established in 1982 by

former US president Jimmy Carter, announced that Britain has committed to contribute £20 million (US\$31 million) to the effort over four years — about one-third of what the centre says is needed — if other donors step up. Better case surveillance and safer water supplies are needed to wipe out the parasitic disease. The condition was once endemic in parts of Africa and Asia but fewer than 1,800 cases were reported in 2010, nearly all of them in the area that is now South Sudan.

Spanish funding

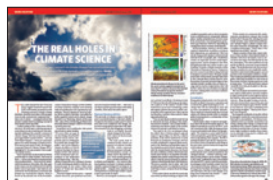
Spain's ministry of science has named eight research centres that will each receive an extra €5 million (US\$6.8 million) in funding over the next five years — the first fruits

of a programme set up by science minister Cristina Garmendia to preserve scientific excellence at a time of economic crisis. Thirty-two further centres are supposed to receive similar funding over the next four years, but only if the programme survives a possible change of government after elections in November. See go.nature.com/2pvrph for more.

Paying for eggs

Women who donate their unfertilized eggs to research deserve to be paid, concluded a report released on 10 October by the influential Nuffield Council on Bioethics, an independent charity based in London. In Britain, direct payment for unfertilized eggs is prohibited; in the United

JOURNALISM AWARD



Nature reporter Quirin Schiermeier last week won the American Meteorological Society Award for Distinguished Science Journalism in the Atmospheric and Related Sciences for his feature 'The real holes in climate science' (*Nature* **463**, 284–287; 2010).

D. CALVERT/AP
States, rules vary by state. The report called for a pilot scheme to look into paying women who donate their eggs for research. It also explored issues involving compensating organ donors and their families. See go.nature.com/rrvczc for more.

BUSINESS

The Scientist retires

The Scientist, a loved and respected US professional magazine for life-sciences researchers, closed down abruptly last week, after celebrating its 25th anniversary. Launched in 1986 by Eugene Garfield, who also founded Thomson Reuters' Science Citation Index, *The Scientist* was most recently bought by the Science Navigation Group in London, which is chaired by Vitek Tracz, founder of open-access publisher Biomed Central. But the magazine's owners could not find a way to make it financially viable. See go.nature.com/9nnrj6 for more.

PEOPLE

XMRV row drags on

The scientist behind a partially retracted paper that linked chronic fatigue syndrome to the XMRV virus has lost her job and is now facing accusations that she has misrepresented data. Judy Mikovits (**pictured**), former



director of research at the Whittemore Peterson Institute for Neuro-Immune Disease in Reno, Nevada, was fired on 29 September after she clashed with the institute's president and co-founder, Annette Whittemore, over the work of another researcher. The following day, Abbie Smith — a virology graduate student at the University of Oklahoma in Norman — said that Mikovits had used an image from the paper to illustrate different results in a presentation last month. See go.nature.com/w2kskt for more.

Chemistry Nobel

Dan Shechtman of the Technion Israel Institute of Technology in Haifa has won the 2011 Nobel Prize in Chemistry for his 1982 discovery of quasicrystals — materials with ordered atomic structures that, unlike those of conventional crystals, never quite repeat themselves. See page 165 for more.

RESEARCH

Climate coalition

Leading climate research organizations from eight European nations have agreed to join forces. The European Climate Research Alliance, a loose coalition launched on 4 October at the European Parliament in Brussels, will focus on research on Arctic climate variability, hydrological cycles in the Mediterranean region, links between climate change and extreme weather, and sea-level rise. Researchers with the founding institutions will set up collaborative research projects at workshops to be held next year. See go.nature.com/qiq86t for more.

University rankings

The California Institute of Technology (Caltech) in Pasadena pipped Harvard University in Cambridge, Massachusetts, to the top spot in an annual ranking of the world's top universities. *Times Higher Education* uses scores in areas such as research, citations and teaching to produce its list. But a rival list compiled by education consultants Quacquarelli Symonds in London — who worked with *Times Higher Education* until 2010 — puts Caltech 12th, with Harvard second and the University of Cambridge, UK, on top. See go.nature.com/pwvwmt for more.

COMING UP

17–21 OCTOBER

The General Conference on Weights and Measures meets in Sèvres, near Paris, to discuss how to redefine the kilogram, ampere, kelvin and mole. go.nature.com/jd84hb

20 OCTOBER

The European Space Agency will launch the first two operational spacecraft of Galileo, Europe's new global positioning system. go.nature.com/wiz7ks

Cosmic missions

The European Space Agency has approved the first two space missions under its Cosmic Vision 2015–2025 programme. In the Solar Orbiter mission, a spacecraft launched in 2017 will venture within 42 million kilometres of the Sun, sampling the solar wind and other ejections of material. The Euclid space telescope will launch in 2019, and will map the structure of the Universe to pin down parameters relating to dark energy and dark matter. Both missions were chosen on 4 October by the agency's science programme committee.

SuperB accelerator

Italy's SuperB particle accelerator, which will be built at the University of Rome Tor Vergata by 2017, was officially launched on Friday. The country's National Institute of Nuclear Physics signed an agreement with the university to set up a lab to oversee the accelerator's construction and operation. SuperB doesn't yet have enough funding to meet its estimated €450-million (US\$610-million) to €600-million price tag. See go.nature.com/xt0000 for more.

► **NATURE.COM**

For daily news updates see:

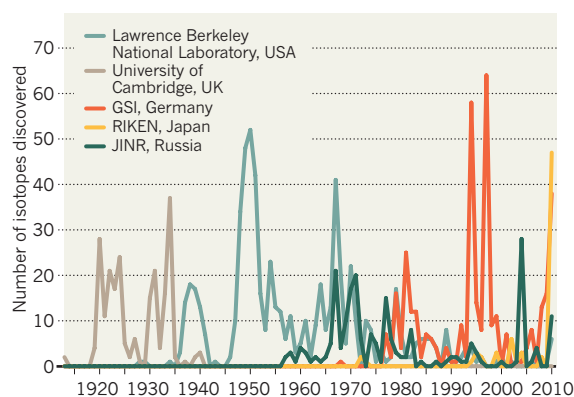
www.nature.com/news

TREND WATCH

A provisional database of the discoveries of almost 3,100 nuclear isotopes shows how the Lawrence Berkeley National Laboratory in California lost ground to other labs during the 1980s when it failed to upgrade equipment. Most finds now come from the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany, the RIKEN Radioactive Isotope Beam Factory in Wako, Japan, and the Joint Institute for Nuclear Research (JINR) in Dubna, Russia. See go.nature.com/viznae for more.

ISOTOPE RANKING REVEALS LEADING LABS

Laboratories in Russia, Germany and Japan have taken the lead from the United States in finding new isotopes.



NEWS IN FOCUS

NOBEL PRIZES Chemistry winner grasped importance of strange results **p.165**

ASTRONOMY Giant solar telescope faces moment of truth **p.166**

AGRICULTURE Brazil cooks up high-tech pinto bean **p.168**

ETHICS France's researchers chafe at new conflict of interest law **p.169**



ROCKEFELLER UNIV.

hopes on the cells that had been his life's work. Together with collaborators around the world, he designed therapies that made use of his own dendritic cells.

"He was running an experiment on himself and was willing to help out with every kind of study. He wanted to help himself, but he also viewed it as an incredible opportunity to learn something," says Ira Mellman, who worked with Steinman to develop his treatments and is vice-president of oncology research at the biotechnology firm Genentech in South San Francisco, California.

On 3 October, Steinman shared the Nobel Prize in Physiology or Medicine for his work, but he never heard the news. At the age of 68, after a four-and-a-half year battle with cancer, he died three days before the award was announced (see *Nature* **478**, 13–14; 2011).

I first met Steinman during my two-year tenure as a science writer in the Rockefeller communications department. I was new to the immunology beat, and he kindly and patiently talked me through the intricacies of dendritic cells and their vast potential. When word of his cancer diagnosis emerged, his students and postdocs talked about it in hushed tones, telling me that immunologists at Rockefeller and beyond were using Steinman's dendritic cells in a personalized immunotherapy. I vaguely pictured his colleagues injecting him with homegrown cells right there in his lab. I could not have been more wrong.

"Everybody around the world who had something to share came forward, and he analysed and chose what looked most promising," says Sarah Schlesinger, a physician-researcher at Rockefeller who worked closely with Steinman and oversaw many of his experimental treatments. "We worked with dozens of colleagues, who helped in designing his therapy, evaluating the tumour and evaluating his immune response, and many worked with us to create single-patient protocols to treat him with experimental immunotherapy that went through the FDA [US Food and Drug Administration]."

Researchers across the field were eager to help the man who had always been generous with his time and knowledge. "Ralph was a collaborator, a competitor, but before everything he was a friend," says Jacques Banchereau, who began working with Steinman in the early 1990s and is now head of inflammation and virology at Roche in Nutley, New Jersey. ▶

Ralph Steinman used his findings to help design treatments that may have prolonged his life.

IMMUNOTHERAPY

A fight for life that united a field

Nobel-prizewinner Ralph Steinman tried to beat his cancer with vaccines based on the dendritic cells he discovered.

BY LAUREN GRAVITZ

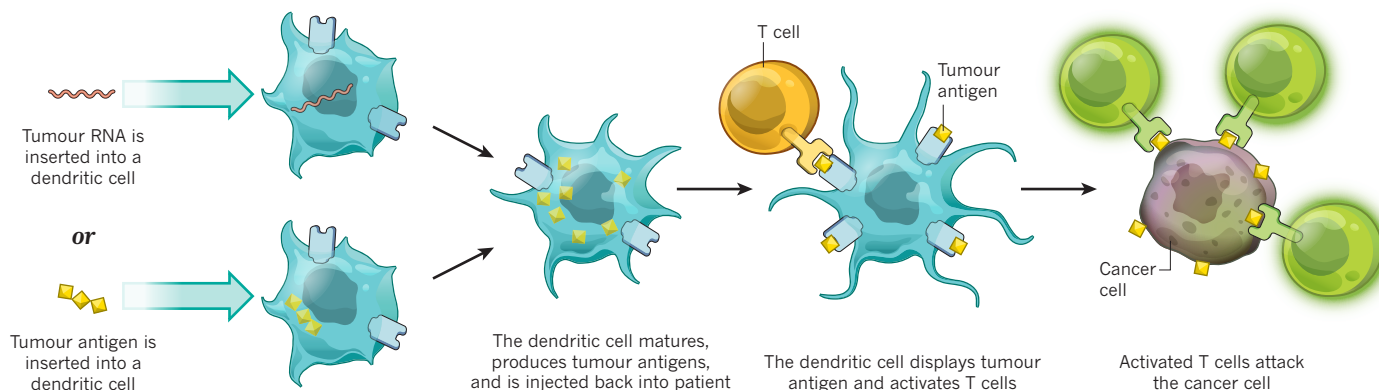
After a young Ralph Steinman co-discovered a new type of immune cell in 1973, he spent years battling to prove its importance in defending the body against pathogens, and to show how it could be used to fight disease. Thirty-four years later, he would look to that same cell to try to save his life.

Dendritic cells — named for their tree-like

branches — direct and regulate the body's immune system by programming other cells to recognize and destroy intruders. Steinman, a physician-scientist at The Rockefeller University in New York, set his sights on using the cells in vaccines to prevent chronic infections, such as HIV and tuberculosis, and in cancer therapies. So when he was diagnosed with advanced pancreatic cancer in March 2007, it was only natural that he would pin his

HOW THE CANCER VACCINES WORK

In the immune system, dendritic cells can be used to trigger a response against cancer cells in various ways. In these two examples, antigens are introduced directly into the dendritic cells or generated *in situ* using tumour RNA. The cells then produce tumour antigens and direct T cells to seek out and destroy the cancer cells.



► Shortly after Steinman was diagnosed he met with two former members of his lab, both of whom now run successful immunotherapy research programmes of their own. Michel Nussenzweig, at Rockefeller, and Genentech's Mellman sat down with him to discuss his case. "It was the weirdest experience, like we were having a lab meeting from the old days: talking about what experiments to do, what needed to be found out, how interesting it was, what you can and can't do," Mellman says. "It was a totally natural scientific discussion, except we were talking about his tumour."

ATTACKING THE CANCER

The scientists hatched a plan: Nussenzweig would take some of the tumour that would be removed during surgery and grow it in mice so that they would have enough material to test. Mellman would start a cell line that could be used to test the tumour's susceptibility to various drugs. A colleague in Toronto would sequence the tumour's DNA in the hope of finding a mutation that could inform drug choices, and a collaborator in Germany would extract peptides from the tumour's surface so that they could later be synthesized for use in a vaccine. And for treatment, Steinman would undergo conventional chemotherapy in combination with as many experimental therapies as made sense for his disease.

Steinman tried eight experimental treatments in all, one at a time, and for each one a single-patient, compassionate-use protocol was submitted to and approved by the FDA. One of the reasons that cancer is so lethal is that it uses a variety of tactics to prevent immune cells from recognizing and attacking it. So Steinman's line-up included three vaccines, all of which were based on dendritic-cell science. The first was GVAX, which has been tested against prostate and other cancers and is being developed at BioSante Pharmaceuticals in Lincolnshire, Illinois. It used irradiated cells from Steinman's tumour that had been engineered to produce a potent stimulator that attracts

and activates dendritic and other immune cells. The second, developed by Argos Therapeutics of Durham, North Carolina, loaded Steinman's dendritic cells with RNA that had been extracted from the tumour (see 'How the cancer vaccines work'). The resulting dendritic cells, now decorated with tumour-specific antigens on their surfaces, were injected back into Steinman to teach other immune cells what to seek and destroy. The third vaccine, which was in clinical trials for melanoma at the Baylor Institute for Immunology Research in Dallas, Texas, attempted to boost the immune response by loading up Steinman's dendritic cells with peptide antigens from the surface of the tumour. Once injected back into the body, the cells would be prepared to recognize the cancer and coordinate an attack.

"It was the ultimate experience in personalized medicine," says Jedd Wolchok, a medical oncologist at Memorial Sloan-Kettering Cancer Center in New York, and one of Steinman's collaborators. It was never quite as tailored as Steinman would have liked, however. He

"It was the ultimate experience in personalized medicine."

believed, and others in the field agree, that the vaccines would probably work better if combined with a checkpoint-blockade inhibitor, which allows cancer-suppressed immune responses to ramp back up to full strength. Together, Steinman believed, the therapies would act as a one-two punch. He tried one such inhibitor, ipilimumab, on its own, but never in combination. Investigators, manufacturers and the FDA must all agree to protocols that involve experimental combinations, and Steinman and his team never got the permission they needed. Ipilimumab received FDA approval for melanoma treatment in March, and trials testing it in combination with dendritic-cell vaccines are only just beginning, in melanoma and renal-cell carcinoma.

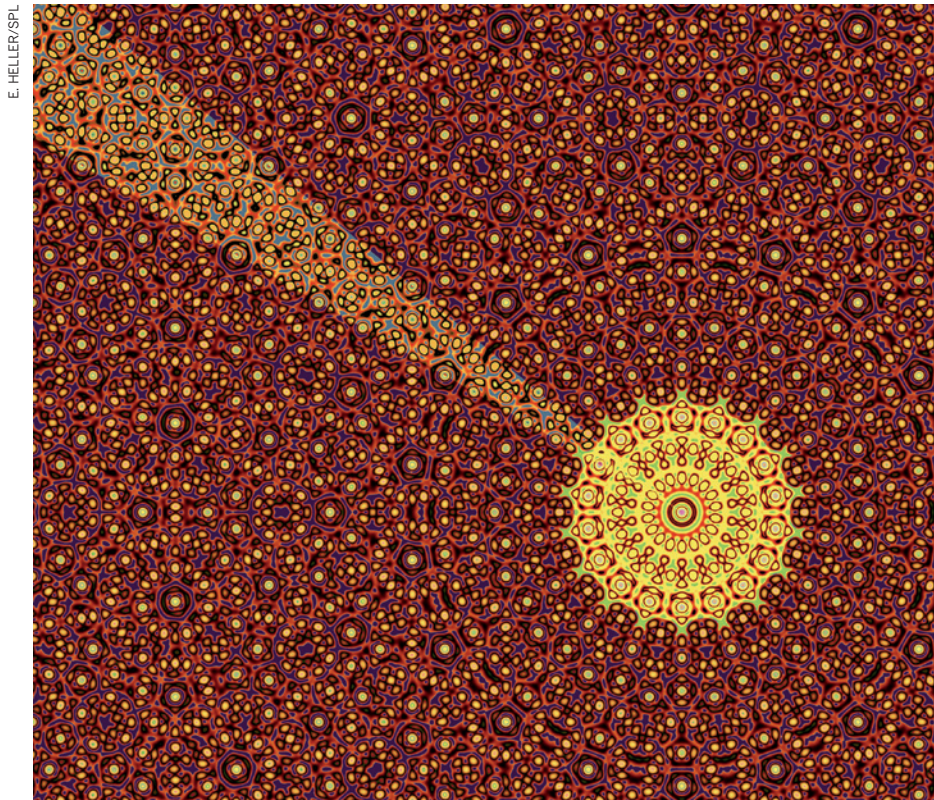
Which, if any, of the therapies extended

Steinman's life is impossible to know. He lived years longer than his initial prognosis — survival time for patients with that degree of pancreatic cancer is usually measured in weeks to months. "Ralph was committed to the idea that his dendritic cells extended his life," Schlesinger says. "Certainly something did, but I don't think we'll ever know for sure what."

Those who monitored his treatment say that he was particularly responsive to a conventional chemotherapy, gemcitabine, which most people with pancreatic cancer become resistant to after just one or two treatment cycles. They also know that he had a measurable immune response: 8% of his cytotoxic, or killer, T cells were targeted to his pancreatic tumour. It was unclear, however, whether that was because of the dendritic-cell vaccines he received or because he had a natural immunity. Nor did anyone know whether the immune response helped to make the cancer susceptible to gemcitabine. "We knew at the outset that we wouldn't be able to tell which therapy made the difference," Schlesinger says. "We only had one patient, so there's no statistical significance."

A controlled experiment it was not, but Steinman's one-man trial moved the field forward. It answered a major question by showing that conventional chemotherapy could be given in conjunction with dendritic-cell vaccines. It bolstered Steinman's contention that experimental therapies should be tested in humans as quickly as possible, owing to the limitations of animal models. And it united the best minds in the field, all fighting for a common cause.

Anna Karolina Palucka, one of the investigators who oversaw the development of Steinman's dendritic-cell vaccine at Baylor, says that she and her colleagues are developing a full programme of immunotherapy against pancreatic cancer using the data gathered from Steinman's solo trial. And, in honour of him, the university will be opening a Ralph Steinman Center for Cancer Vaccines. "He started a completely new field," Palucka says. ■



A quasicrystal's atomic structure never quite repeats itself, as shown in this computer simulation.

NOBEL PRIZE

Persistence pays off for crystal chemist

Discovery of mosaic material wins Nobel prize.

BY RICHARD VAN NOORDEN

As Louis Pasteur famously pointed out: “In the fields of observation, chance favours only the prepared mind.” Stumbling across a curiosity is not enough — understanding it, and convincing the world of its importance, are key.

Dan Shechtman at the Technion Israel Institute of Technology in Haifa knows this only too well. Last week he was awarded the Nobel Prize in Chemistry for his 1982 observations of quasicrystals: materials with a mosaic-like atomic array that never quite repeats, thus flouting the established rules of crystal structure (see ‘Chemistry’s lone heroes’).

Yet Shechtman was not the first to spot evidence of the crystals. “Given the relative simplicity of making these materials, it is certain that they would have been seen by numerous scientists before, who dismissed them because they didn’t fit the rigid rules of crystallography,” says Veit Elser, a physicist at

Cornell University in Ithaca, New York.

Indeed, in 1979, Marc van Sande, a 27-year-old doctoral student working in the Electron Microscopy for Materials Science (EMAT) group at the University of Antwerp, Belgium, had recorded electron diffraction patterns from metal alloys that showed clear evidence of quasicrystals. Close to the end of his PhD, van Sande just filed the confusing patterns in the EMAT library: he was keen to take up his new job with the Belgian materials-technology company Umicore, where he is now an executive vice-president.

“We were making so many new discoveries every week with the high-resolution electron microscopes that the more awkward things were set aside,” recalls van Sande. “I’m realistic about it: seeing the pattern is a long way from investigating it and publishing it.”

Shechtman, unaware of van Sande’s work, observed similar odd diffraction patterns about three years later — and grasped their importance. He had the self-belief to

plough on with his work, despite the scorn of luminaries such as chemist and two-time Nobel prizewinner Linus Pauling. “If you have repeated your observations and are sure you are correct, then listen to others but don’t give up because people tell you ‘this cannot be,’” Shechtman told *Nature*.

Thirty years ago, scientists were taught that all crystalline materials were composed of atoms packed into regularly repeating three-dimensional lattices, such as the hexagonal honeycomb of a beehive. This definition dictated that the lattice must have basic repeating units with particular symmetries: these units could be rotated by one-half, one-quarter or one-sixth of a full circle and still look the same. Pentagonal symmetry was ruled out, because no perfectly regular lattice could exhibit it.

On 8 April 1982, Shechtman, who was on sabbatical at the US National Bureau of Standards (now the National Institute of Standards and Technology, NIST) in Gaithersburg, Maryland, found that an artificial alloy of aluminium and manganese disobeyed the rules.

When he shot electrons through the material, they created a regular diffraction pattern, apparently proving that the material’s atomic structure consisted of orderly repeating elements. But that pattern showed a forbidden symmetry — it could be rotated by both one-tenth and one-fifth of a full circle and would still look the same. In his laboratory notebook, Shechtman wrote: “10 Fold???”

Others did their best to persuade him that his discovery was wrong. “People didn’t believe me,” he says, adding that his dogged pursuit of the problem even led his group director to suggest he move to another team. Shechtman finally got his findings published in November 1984, along with Ilan Blech, a materials scientist at Technion; John Cahn, a physicist at NIST; and Denis Gratias, a crystallographer then at the Centre for Metallurgic Chemistry in Vitry, France (D. Shechtman *et al. Phys. Rev. Lett.* **53**, 1951–1953; 1984).

By chance, mathematicians Paul Steinhardt and Dov Levine — both then at the University of Pennsylvania in Philadelphia — were at the same time completing a rigorous theory

of the three-dimensional version of mathematical curios known as Penrose tilings, structures with an apparent five-fold symmetry that were created by British mathematician Roger Penrose in the 1970s. Steinhardt coined the term ‘quasicrystals’ for the resulting structures: neither classically ▶



“People didn’t believe me.”

Dan Shechtman

► periodic crystals, nor a glass-like mess of disordered atoms. They were what Shechtman had seen in his metallic alloy.

Other examples soon flooded in from around the globe. In 2009, Steinhardt and other researchers reported the first quasicrystal structure to be seen in a natural material: an alloy of aluminium, copper and iron reported to have come from 200-million-year-old rocks in Russia's Koryak Mountains (L. Bindi *et al. Science* **324**, 1306–1309; 2009).

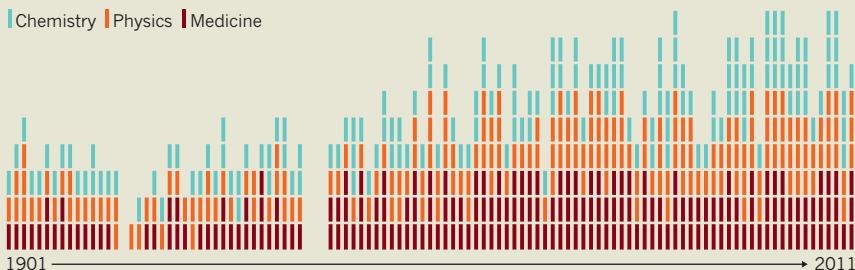
It still isn't clear how atoms assemble into quasicrystal structures, and the discovery has found few real-world applications. However, quasicrystals do have potential: they are very hard, are poor at conducting heat and electricity, and have non-stick surfaces. But Shechtman's key contribution to chemistry lies in opening scientists' eyes to the possibility of new forms of matter, notes Sven Lidin, an inorganic chemist at Stockholm University and a member of the Nobel Committee for Chemistry. As Lidin wrote in his description of the award: "The discovery of quasicrystals has taught us humility." ■

CHEMISTRY'S LONE HEROES

How many prizewinners does it take...

Dan Shechtman got the chemistry Nobel prize all to himself this year — something that has happened in 61% of all chemistry Nobels, but only in 45% of Nobels in physics and 37% in medicine. As large teams and international scientific collaborations come to dominate research, the Royal Swedish Academy of Sciences

is likely to find itself bumping up against the rule that the Nobel prize can be awarded to no more than three people in any one discipline — and the strain is already evident (see chart). Since 2000, the three scientific Nobels have been won by the maximum possible number of nine researchers on five occasions.



ASTRONOMY

Sun-watchers hope giant telescope will get green light

Observatory would reveal structures that trigger sunspots and space weather.

BY ERIC HAND

Close and bright though it is, the Sun still defies a thorough understanding. One reason is that some of the features on its roiling surface are too small and short-lived to be studied even by the world's largest solar telescopes.

That will change if the US National Solar Observatory (NSO) proceeds with its latest project — the Advanced Technology Solar Telescope (ATST), a Sun-gazing behemoth due to be built on the summit of Haleakala, the highest peak on the Hawaiian island of Maui. This month, an officially appointed arbiter will weigh the scientists' goals against objections raised by conservationists and Native Hawaiian groups to decide whether the US\$298-million project can break ground later this year.

With more than twice the aperture of existing solar telescopes (see 'Eyes on the Sun'), the 4-metre ATST will be large enough to tease out small structures on the Sun, particularly magnetic flux tubes — the hitherto unseen precursors to sunspots. Sunspots, in turn, give

rise to giant coronal loops and flares, which can unleash bursts of radiation and cause magnetic disturbances that sometimes threaten spacecraft, communication networks and power grids.

Heliophysicists say the improved resolution should help solar science to move into the challenging terrain of predicting space weather. "The societal need is really driving this community," says Thomas Zurbuchen, a solar physicist at the University of Michigan in Ann Arbor, who is vice-chair of a decadal survey for heliophysics that is due to release its research priorities next March.

Although not the only phenomenon the ATST will explore, flux tubes are of especial interest because these short-lived, 100-kilometre-scale features are thought to be conduits of magnetic field lines through the solar surface. Studying them as they evolve should shed light on larger-scale solar phenomena that can affect Earth. "If you want to understand coronal

heating and solar-wind acceleration, then you have to understand these fundamental scales," says Todd Hoeksema, a solar physicist at Stanford University in California.

Ultimately, solar astronomers hope to monitor the Sun 24 hours a day using a global network of telescopes. A European telescope of similar size is planned for the Canary Islands, and India is considering a 2-metre instrument. But the ATST will be built first if it passes a final hurdle.

Last December, more than seven years after the NSO chose the site, Hawaii's Board of Land and Natural Resources gave permission to develop it. A group called Kilakila O Haleakala ('Majestic is Haleakala' in Hawaiian) has contested the decision. An endangered seabird, the Hawaiian petrel or 'ua'u (*Pterodroma sandwichensis*), nests near the proposed site. Furthermore, some Native Hawaiians say that the telescope's stark white enclosure — necessary to control heat-induced air currents within the scope's optical path — will scar a sacred area. But the telescope builders say they will do all they can to mitigate the impacts. Construction

"Societal need is really driving this community."



The summit of Haleakala in Hawaii will gain its largest telescope yet if plans are approved.

workers will limit vibrations that could collapse the petrels' burrows, and will receive 'sense of place' training to avoid culturally insensitive missteps.

Honolulu-based lawyer Steven Jacobson, the arbiter appointed by the board to re-evaluate the permit, says that he will hand in his recommendation in the next week. NSO director Stephen Keil is cautiously optimistic that Jacobson will give the telescope the green light — although he has seen the process take plenty of detours before. "It keeps me awake every night," he says. "This is part of doing business in Hawaii."

There are other issues to be resolved — not least whether the cash-strapped National Science Foundation can find enough money. A \$146-million infusion from the 2009 stimulus-package fund has paid for early work on the telescope's mirror. But Keil says the project will need "make up" funding of around \$30 million a year — six times the amount it received in 2011 — if it is to begin operations by 2018 as planned.

By then, a new solar clearing house will be ready to receive data from the telescope. On 30 September, the board of the Association of Universities for Research in Astronomy, which oversees the NSO, announced that the observatory's headquarters are to move to the University of Colorado at Boulder. The move, planned for 2016, will spell the end for the NSO's current homes in Sunspot, New Mexico, and Tucson, Arizona, and for the ageing solar telescopes at the two sites. Administrators hope that the consolidation of staff and resources, coupled with the anticipated power of the ATST, will usher in a new era for the field, allowing astronomers to see the Sun in an entirely new light. ■

EYES ON THE SUN

The largest current and proposed solar telescopes.

MCMATH-PIERCE TELESCOPE

APERTURE: 1.6 metres
LOCATION: Kitt Peak, Arizona
FIRST LIGHT: 1962

SWEDISH SOLAR TELESCOPE

APERTURE: 1 metre
LOCATION: La Palma, Spain
FIRST LIGHT: 2002

NEW SOLAR TELESCOPE

APERTURE: 1.6 metres
LOCATION: Big Bear, California
FIRST LIGHT: 2009

GREGOR

APERTURE: 1.5 metres
LOCATION: Tenerife, Spain
FIRST LIGHT: 2012 (expected)

ATST

APERTURE: 4 metres
LOCATION: Haleakala, Hawaii
FIRST LIGHT: Proposed

NATIONAL LARGE SOLAR TELESCOPE

APERTURE: 2 metres
LOCATION: Ladakh, India
FIRST LIGHT: Proposed

EUROPEAN SOLAR TELESCOPE

APERTURE: 4 metres
LOCATION: Canary Islands, Spain
FIRST LIGHT: Proposed



Pinto beans, a Brazilian staple, could soon be resistant to the devastating golden mosaic virus.

AGRICULTURAL SCIENCE

Brazil cooks up transgenic bean

Approval draws criticism over transparency and safety tests.

BY JEFF TOLLEFSON

Paired with rice or steeped in feijoada stew, beans are an essential feature of Brazilian cuisine. So great is Brazil's love of legumes that demand often outstrips domestic supply, forcing the country to import beans from Argentina, Bolivia and China. But this relationship could face the ultimate test as Brazilian scientists roll out a transgenic pinto bean (*Phaseolus vulgaris*) engineered to fend off one of the crop's most devastating enemies: the golden mosaic virus.

Approved on 15 September by the Brazilian National Technical Commission on Biosafety (CTNBio), the transgenic bean uses RNA interference to shut down replication of the virus (K. Bonfim *et al. Mol. Plant Microbe Interact.* **20**, 717–726; 2007). A product of more than a decade of home-grown research, the bean could begin appearing on tables across the country as early as 2014.

"It is an extremely important crop for

our small farmers," says Francisco Aragão, a plant geneticist who led the work for the Brazilian Agricultural Research Corporation (EMBRAPA), the research arm of the Ministry of Agriculture, based in Brasília.

The biosafety commission has taken a favourable position towards biotechnology in past years, helping Brazil to become the world's second-largest producer of genetically modified (GM) crops, behind the United States. Farmers have planted vast tracts of GM maize (corn), soya and cotton with little public resistance, but EMBRAPA is now tinkering with a product that people eat in large quantities every day, says Rubens Nodari, a plant geneticist at the Federal University of Santa Catarina in Florianópolis.

Environmental groups and a presidential advisory panel, the National Council for Food Security and Nutrition, have called for more

transparency in biotechnology science and decision-making, and increased research to rule out health risks stemming from the bean. Nodari, a former member of CTNBio who has long questioned transgenic crops, says that the commission improperly granted EMBRAPA's request for confidentiality regarding key aspects of the genetic engineering. "We don't know what we will be eating tomorrow in Brazil," he says.

Current members of the commission have aggressively defended their decision. In a media interview after the decision last month, Edilson Paiva, president of CTNBio, said that Nodari and other opponents of genetic engineering are taking an ideological position aimed at "promoting fear and uncertainty" as they demand that scientists provide the impossible: guarantees of absolute safety.

EMBRAPA says that it must keep core information about genetic insertions confidential, to allow it to patent the work. The details will help the agency to develop bean varieties that are resistant to the golden mosaic and similar viruses, says Aragão, who is a member of CTNBio but abstained from the decision on the beans.

Aragão notes that safety analyses showed no reason for concern regarding the beans. He says that whereas some other GM crops produce unfamiliar proteins that could in theory cause an allergic reaction when eaten, the GM pinto bean produces only small snippets of RNA, tailored to react with and neutralize RNA from any invading virus. Herve Vanderschuren, a biotechnologist at the Swiss Federal Institute of Technology in Zurich, adds that plants naturally produce similar RNA snippets to defend themselves from viral attack, and there is no evidence that this common molecular warfare is dangerous to humans.

With approval secured, EMBRAPA must now conduct a further round of field trials to ensure that the transgenic bean produces yields comparable to those of existing varieties. Aragão hopes that the strain will not only boost yields, but also enable planting on as much as 200,000 hectares of land on which the golden mosaic virus is so prevalent that farmers cannot grow beans at all at present. Brazil produces some 3.5 million tonnes of beans per year already, and Aragão says that the transgenic bean could increase production by 10–20%, enough to offset imports and soften the price spikes that accompany domestic shortages.

"The best part of this story is that the bean was developed in Brazil for the Brazilian farmers," says Vanderschuren, who is part of a consortium working with researchers in Kenya, Tanzania and South Africa to apply the same technology to local crops, including cassava.

EMBRAPA is already looking to develop other virus-resistant beans, including common black beans and the popular carioca bean. "It's very easy to transfer this gene to any other variety," says Aragão. "That's the next step." ■

➔ **NATURE.COM**

For more on how science can increase food production: go.nature.com/vt3rqo

ETHICS

France toughens conflict rules

Law demands greater transparency for health experts.

BY DECLAN BUTLER

A draconian new law aims to toughen France's relaxed approach to conflicts of interest for scientists who advise the government on pharmaceuticals. Approved last week by the French National Assembly, the law is intended to restore public confidence in the country's embattled drug regulatory system. It mandates fines of up to €30,000 (US\$40,000) for advisers who fail to declare all their competing interests.

The law, which is yet to be voted on by the French Senate, is a response to the 'Mediator scandal', in which the French authorities are alleged to have acted too slowly when they withdrew the diabetes drug Mediator (benfluorex) from the market in November 2009. Press reports have suggested that advisers' ties to the pharmaceutical industry may have contributed to a tardy response to long-standing concerns about potentially fatal cardiac side effects. Some scientists argue, however, that the Mediator affair has been blown out of proportion by the media, and that the facts of the case — including what, if any, role conflicts of interest had — will become clear only with the conclusions of an ongoing judicial investigation.

Given the outcry, it is reasonable for the government to try to rebuild public trust in the drug regulatory system, says Claude Le Pen, a health economist at Paris-Dauphine University. But he and other experts argue that the law is largely a short-term, populist response that fails to address more fundamental weaknesses in the regulatory system. "The law's aim is more political, more a public-communication exercise, than a reform of the system," he says.

The national drug regulator, AFSSAPS, and many other French health bodies have long had policies requiring experts to declare any interests that might raise questions about impartiality, such as being a paid consultant for a particular drug company. But these have not been well enforced and are often ignored.

The new law brings in a harmonized national

system for handling experts' competing interests throughout the health system, and gives it penal teeth, says Marc Rodwin, an expert in comparative health law at Suffolk University Law School in Boston, Massachusetts.

Under the law, external experts and in-house staff within the health system would have to declare all competing interests. Companies would also be obliged to maintain a public register of all agreements with, and payments to, anyone involved in health care. This provision, which experts believe is far more sweeping than in other countries, goes beyond expert advisers,



Some believe that conflicts of interest may have slowed the decision to withdraw Mediator.

and applies to all medical professionals, associations, charities, institutions and students, as well as the trade press and learned societies.

Minutes from committee meetings within the regulatory system would in future be publicly available, with the exception of discussions involving confidential personal-health or commercial information. The law would also dissolve AFSSAPS and replace it with a National Agency for Medicine Safety, although whether this will be any more than a symbolic change is unclear.

But the law fails to address the regulatory agencies' heavy reliance on external academic experts, many of whom inevitably have ties with industry. Rodwin argues that the agencies have neglected to build up a sufficient corps of in-house experts, as the US Food and Drug Administration and many other regulatory

agencies have done. Le Pen adds that the new law's sanctions could discourage researchers from serving on government advisory committees — generally an unpaid role.

In most countries, ties with industry are widely accepted, and transparency is seen as the best way to deal with potential conflicts. In France, however, honest scientists who declare competing interests are often not seen by the public as transparent; rather, they are frequently vilified in the French media and Internet forums as lackeys of industry, Le Pen says.

That attitude is deeply rooted in the French psyche. In contrast to the United States or Germany, people in France — paradoxically, another leading capitalist economy — tend to be suspicious of profit, entrepreneurship and industry, says Francis Megerlin, an expert in comparative law at the University of Paris-Descartes, currently at the Berkeley Center for Health Technology in Oakland, California.

The real issue, he says, is that the government is trying to put in place legislation loosely modelled on that set out in the US Physician Payments Sunshine Act of 2009, which requires drug companies to disclose payments made to physicians. "The problem is that on the sociological and ideological level the context is completely different," says Megerlin. Scientists say they see no easy solution to the dilemma.

The law is being pushed through the legislature using the 'accelerated procedure', which means that it requires only one reading in the parliament and the Senate, instead of the usual two. The text will come before the Senate on 26 October, but observers expect it to fly through with few significant changes. ■

CORRECTION

The Editorial 'Heritage threats' (*Nature* **477**, 510; 2011) originally stated that Ireland was leading the European Joint Programming Initiative. In fact, Italy has overall responsibility for coordination.


**MORE
ONLINE**

TOP STORY



Safety fears put India's nuclear future on hold
go.nature.com/aabr9w

MORE NEWS

- MERMAIDs detect distant earthquakes go.nature.com/lusa8y
- HIV risk of progesterone contraceptives was long suspected go.nature.com/cwmy5n
- Faster-than-light neutrinos face time trial go.nature.com/lcopy7

PODCAST



Editor Henry Gee talks about Futures, *Nature's* long-running science-fiction column go.nature.com/bozkn5



AFTER THE ICE

As the Arctic thaws, can science help to chart a sustainable path for the north?

Last winter, parts of the Canadian Arctic basked in record-breaking warmth. In the town of Coral Harbour, at the mouth of Hudson Bay, temperatures rose above freezing for a few days in January for the first time ever. Across the Arctic, extreme climate conditions are becoming the norm, even as the region faces other profound changes, such as the growing political power of indigenous peoples and the race to extract mineral resources (see page 172).

This week, *Nature* examines how these changes are affecting scientific access to the north (see page 174), and what scientists should do to keep Arctic development green (see page 179) and peaceful (see page 180). Some are calling for international regulations to safeguard the environment as ship traffic increases (see page 157). Both research and development need to consider the views of local peoples, and scientists are learning how to do so (see page 182). Locals can provide insight into environmental changes; scientists might help them to be heard.

There is a long tradition of international scientific cooperation in the Arctic. That tradition must be preserved and expanded, even as nations in the region push forward with territorial claims under the UN Convention on the Law of the Sea. Countries should work together, perhaps through the Arctic Council, to ensure that researchers from all nations are allowed access to all parts of the Arctic Ocean.

At the same time, scientists should make their data available in public databases as soon as possible after collection. The far north is changing faster than anywhere else on Earth, with potentially vast impacts on climate as carbon-rich permafrost melts, and dark ground and water exposed by the retreating ice soak up more heat from the Sun. It is crucial that science keeps up. ■

KEENPRESS/NATIONAL GEOGRAPHIC STOCK



AFTER THE ICE

Science at the new Arctic frontier
nature.com/arcticfrontier

THE NEW NORTH

The Arctic covers around 5% of the planet's surface, but it is capturing a disproportionate amount of attention. With temperatures rising at twice the global rate, the region's summer sea ice is shrinking rapidly, making access easier than ever before. At the same time, countries are racing to claim parts of the Arctic's sea floor and the vast deposits of hydrocarbons that lie beneath it.

DISAPPEARING SEA ICE

On 9 September, the area of the Arctic covered by sea ice reached a near-record low of 4.33 million square kilometres.

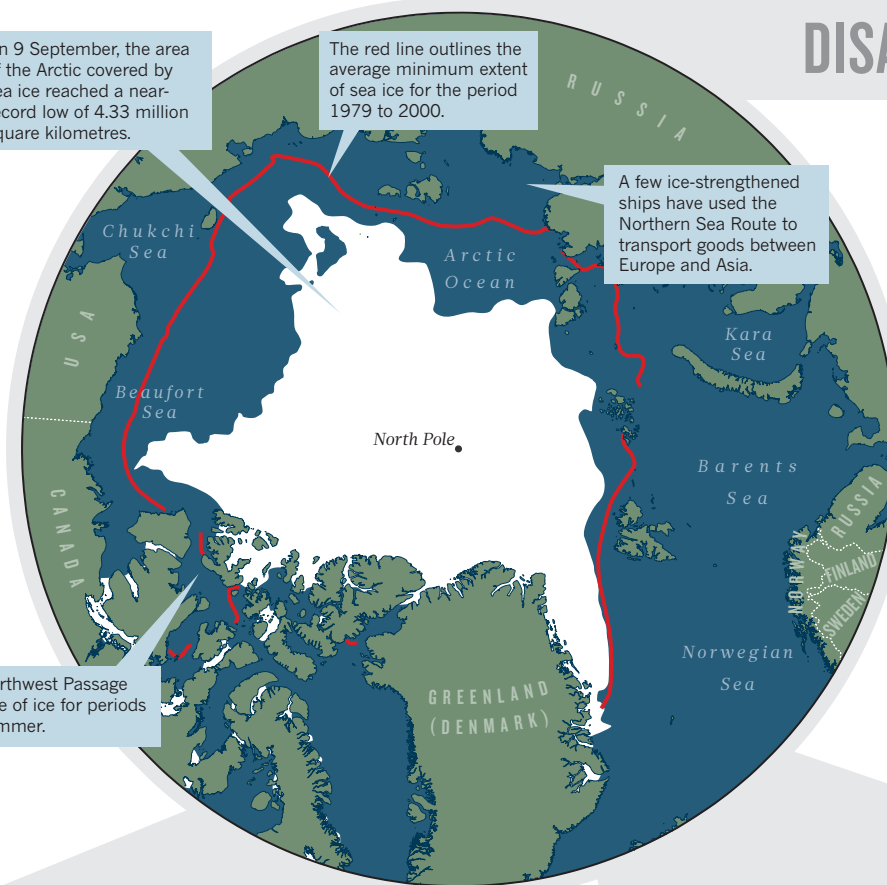
The red line outlines the average minimum extent of sea ice for the period 1979 to 2000.

A few ice-strengthened ships have used the Northern Sea Route to transport goods between Europe and Asia.

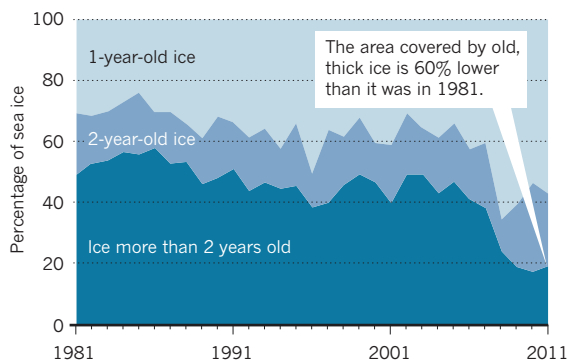
The Northwest Passage was free of ice for periods this summer.

Since satellite observations started in 1979, the September sea-ice extent has declined by 12% per decade, and the past 5 years have marked the lowest on record. The ice cover is thinning (**below, left**), making it more vulnerable to warmer temperatures. Forecasts by climate models (**below, right**) suggest that summer sea ice will largely disappear in the second half of the century, but the current rate of ice loss exceeds the models' forecasts, suggesting that ice-free conditions could arrive sooner.

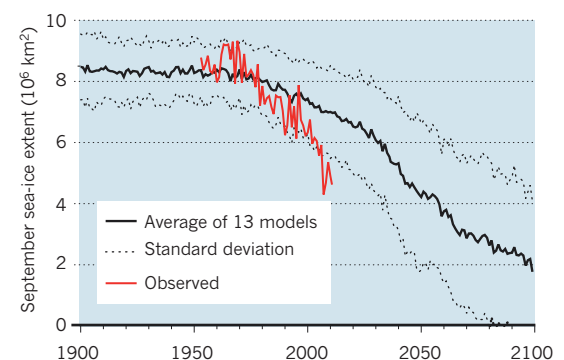
SOURCE: US NATIONAL SNOW AND ICE DATA CENTER



AGE OF ARCTIC SEA ICE



MODELLING ICE LOSS

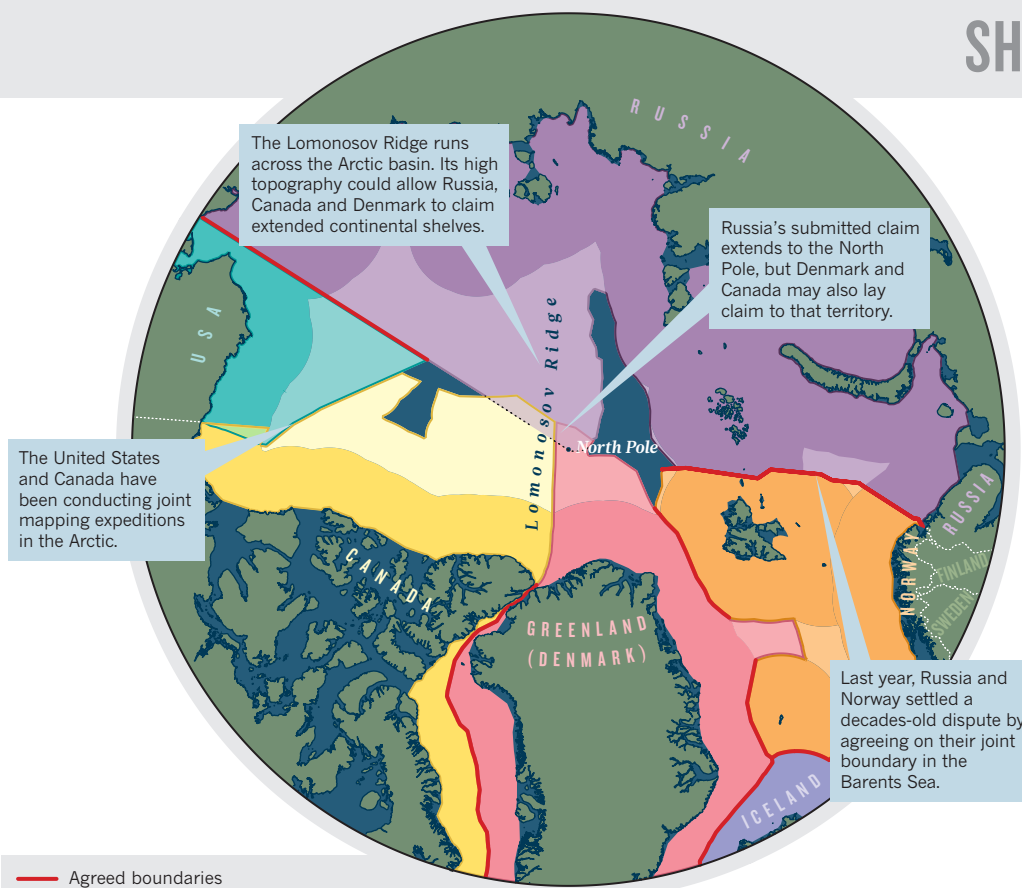


J. MASLANIK

SOURCES: J. STROEVE ET AL. GEOPHYS. RES. LETT. **34**, L09501 (2007); M. HOLLAND

SHIFTING BORDERS

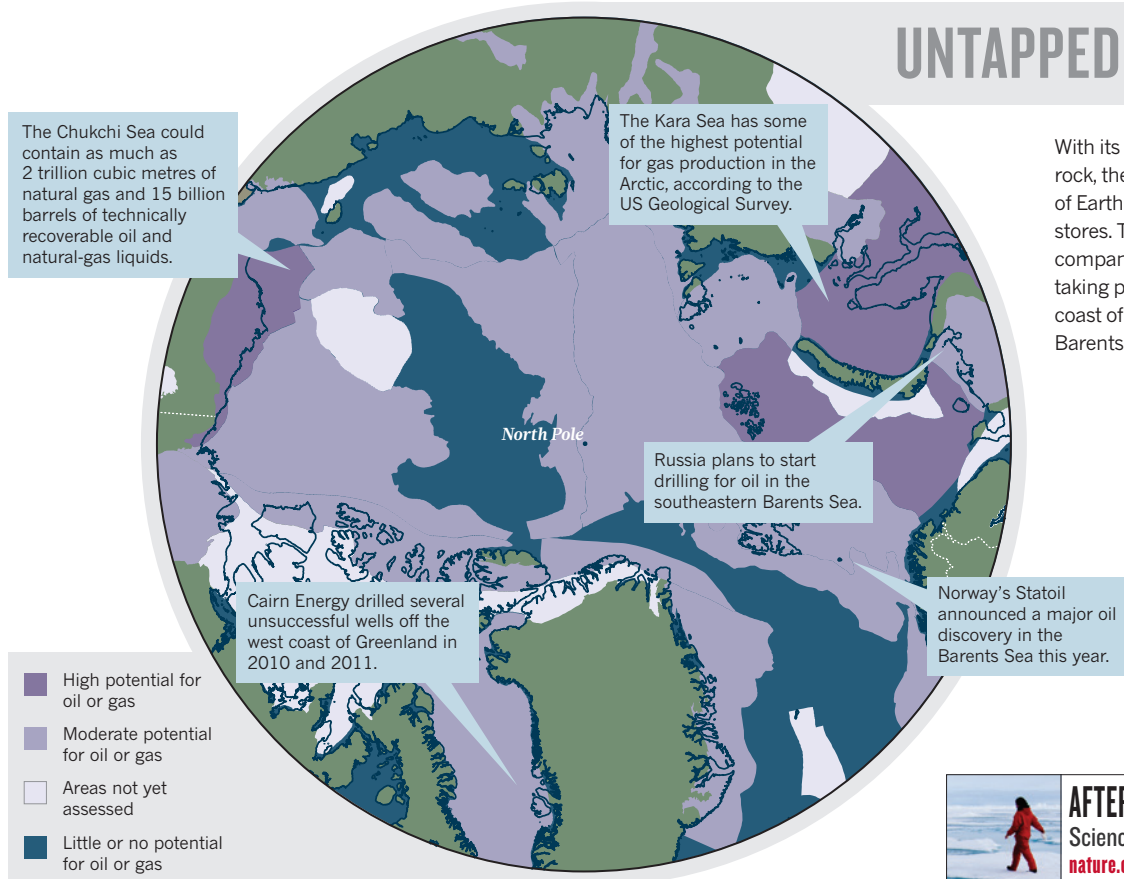
SOURCE: DURHAM UNIVERSITY INTERNATIONAL BOUNDARIES RESEARCH UNIT



Under the United Nations Convention on the Law of the Sea, countries can claim rights to seabed resources in the Arctic Ocean, depending on their coastline and the sea-floor geology. Dark shading on this map represents each nation's existing exclusive economic zone, which extends up to 370 kilometres from its coastline. Lighter shading depicts extended regions to which countries may be eligible. Russia and Norway are the only Arctic nations to have submitted their bids.

USGS

UNTAPPED RESOURCES



With its thick piles of sedimentary rock, the Arctic may hold some of Earth's biggest hydrocarbon stores. The high price of oil is driving companies northwards, with drilling taking place or planned off the coast of Greenland and in the Kara, Barents and Chukchi seas.



AFTER THE ICE

Science at the new Arctic frontier
nature.com/arcticfrontier

OPEN WATER

As the ice melts, fresh obstacles confront Arctic researchers.

BY DANIEL CRESSEY

Last month, US researchers took a 4,000-tonne gamble when they steered the *Marcus G. Langseth* through the Bering Strait and into the Arctic Ocean. The 72-metre research vessel was not built to plow through ice, so it had never ventured that far poleward before.

But the rules are changing quickly in the new north. Managers at the US National Science Foundation (NSF), which owns the ship, decided to send the *Langseth* into the Arctic after reviewing satellite images that showed that the intended survey area in the Chukchi Sea had been largely clear of ice for four of the past five summers.

In an e-mail to *Nature* during the cruise, its principal investigator, Bernard Coakley, said: “We are rolling the dice a bit to take her up north.” But the bet paid off for Coakley, a marine geologist at the University of Alaska Fairbanks. Sea-ice coverage was at near-record lows this summer, and the *Langseth* — due back in dock this week — has not encountered any troubling ice.

With the Arctic warming roughly twice as fast as the rest of the globe, there is more need than ever to monitor the changing conditions there. And the retreating summer sea ice is opening up new options for scientists who want to explore the once difficult-to-reach Arctic waters, allowing them, for example, to use vessels other than icebreakers.

But the scientists are not alone. Businesses, too, are racing to exploit the Arctic — for tourism, fishing, transportation and, especially, resources such as hydrocarbons. According to the US Geological Survey, the Arctic could hold up to 30% of the world’s undiscovered gas and as much as 13% of its undiscovered oil¹.

Governments keen to access this wealth are stepping up their activities in the area as a prelude to claiming rights to resources in vast swathes of territory under the United Nations Convention on the Law of the Sea (UNCLOS). When Russia planted a flag on the sea bed under the North Pole in 2007, many people saw the action as a symbolic statement

about the country’s territorial ambitions — a view bolstered this July when Russia pledged to station two brigades permanently in the Arctic. The next month, Canada launched its annual sovereignty operation in the Arctic and claimed that it now had more military capability in the region than ever before.

By some accounts, all this bluster points to a new cold war that could hamper scientists working in the Arctic. But geologists, oceanographers and others who have been conducting research in the region generally see more cooperation than competition.

“What you read in the media is geopolitical conflict,” says Hajo Eicken, a sea-ice researcher at the University of Alaska Fairbanks. “What we see is quite the contrary. In many cases, you can do Arctic research only if you have good international collaboration in place.”

For researchers, a bigger problem is securing ship time. With the flurry of interest in the Arctic, scientists must compete with drilling companies and others for time on ships designed to operate in the region, which are in short supply. That is how Coakley and his colleagues ended up testing the *Langseth* in the Chukchi Sea.

TERRITORIAL AMBITIONS

When Coakley submitted his proposal to the NSF to investigate the geology of the region, he originally requested the *Healy*, the main US icebreaker, which had been used in past expeditions in the Chukchi. But several factors conspired against that plan, and the *Healy* became overbooked. The icebreaker spent late August and September cruising through the Arctic alongside the Canadian coastguard’s icebreaker *Louis S. St-Laurent*. As part of a multi-year project between the two governments, the ships mapped the extended continental shelf off North America to gather information that may help both countries to stake claims to parts of the Arctic sea floor. ▶



AFTER THE ICE

Science at the new Arctic frontier
nature.com/arcticfrontier



TOP QUESTIONS IN ARCTIC RESEARCH

The Arctic is warming twice as fast the rest of the planet. Here are five issues of particular interest to scientists.

1 ICE

- How are the Greenland ice sheet and other sections of Arctic land ice responding to climatic warming?
- How will the ice respond in the future?
- Are there thresholds beyond which the Greenland ice sheet will become unstable?
- How will melting ice and other freshwater factors alter the Arctic Ocean and the thermohaline circulation that runs through the global ocean basins?

2 AMPLIFICATION

- Why has the Arctic warmed so much faster than the rest of Earth?
- How much of the warming is caused by changes outside the region — such as the influx of heat and pollution from lower latitudes — and how much by local factors, such as regional pollution or feedbacks from the loss of sea ice?
- What is the role of natural variability in the Arctic changes, and how is that likely to evolve?
- How will human activities in the Arctic, such as ship traffic, affect the region?

3 CARBON

- How will changes in the Arctic climate alter the carbon cycle? For example: will melting permafrost release extra methane? Or will the warming boost plant growth, increasing the sequestration of carbon dioxide?

4 ECOSYSTEMS

- How will changes in the Arctic affect ecosystems? This applies not only to the climate but also to shipping, oil and gas drilling, pollution and fishing.
- How well can ecosystems adapt?

5 PEOPLE

- How will Arctic communities, including indigenous peoples, be affected by changes in climate, ecosystems, and the increasing economic activity in the region?



US COAST GUARD PHOTO BY PETTY OFFICER PATRICK KELLEY

The icebreakers *Louis S. St-Laurent* (top) and *Healy* are taking part in a multi-year international Arctic survey to map the Arctic continental shelf.

AUTONOMOUS VEHICLES

New eyes for science

Remote and largely ice-bound, the Arctic is one of the world's most difficult places to study. Icebreakers are in short supply and many satellite-borne sensors are unable to peer through the thick clouds that often shroud the region. So researchers are increasingly using autonomous vehicles, such as underwater craft to collect data below the ice and drones that soar above it.

In the past, Arctic researchers relied on relatively unsophisticated tools, such as buoys and sensors that drift on the pack ice. But autonomous vehicles are much more mobile and can collect different data, from sea-ice thickness to bottom topography.

"We're further along than people realize," says James Maslanik, a climate modeller at the University of Colorado, Boulder, who has flown unmanned aerial vehicles, or UAVs, over thousands of kilometres of ocean and ice. The planes are used mainly "to fill in coverage that the big, manned research aircraft are not suitable for", he says.

Maslanik says that teams from Norway, Russia, Denmark and the United Kingdom have conducted polar work with various UAVs, ranging from small 'hobby aircraft' planes fitted with autopilots to high-tech drones such as NASA's Global Hawk.

Researchers are also testing autonomous vehicles that can dive through openings in the ice and then navigate back to the same holes. In 2007, a team from Woods Hole Oceanographic Institution in Massachusetts used a pair of 250-kilogram submersibles to study the sea floor more than 4,000 metres below the surface, where they discovered a previously undocumented form of deep-sea volcanic eruption⁴. Earlier that year, another team used a smaller device to map the underside of the sea ice. **D.C.**



NASA's aerial drone Sierra was used to study sea ice in the Fram Strait.

► Under the UNCLOS, a nation can claim rights to the seabed beyond the usual 200-nautical-mile zone of control — known as an exclusive economic zone — if it can prove that the claimed region is a natural extension of the country's continental shelf. That requires extensive mapping expeditions, which have generated a large amount of work for marine geologists in recent years, but have also sucked up precious icebreaker time.

The United States has not yet ratified the UNCLOS, and signatories have 10 years after they ratify to make formal claims. So it remains unclear exactly what areas might be snapped up in this process. Some projections suggest that the nations bordering the Arctic Ocean will submit claims that cover the vast majority of that region.

Interest in the Arctic extends far beyond its neighbours. China and South Korea, for example, have built large icebreakers to maintain their presence in the north.

Commercial concerns are driving much of the activity in the Arctic Ocean, with oil companies, in particular, jockeying for position. Exxon, for example, beat out BP this summer by signing a multi-billion-dollar agreement with Russian petroleum giant Rosneft to drill jointly in the Kara Sea. Cairn Energy in Edinburgh, UK, is sinking exploratory wells off the west coast of Greenland, and Shell is moving forward with plans to drill in the Chukchi Sea off Alaska next year.

At the same time, tourist ships are increasingly plying the waters off the west coast of Greenland. Between 2000 and 2010, the number of cruise ships visiting the island more than trebled, as did the number of tourists on those ships, according to the Greenland tourism and business council. The north has become a popular destination.

ICE-FREE FUTURE

Opportunities for travel in the Arctic will expand as its sea-ice cover continues to wither. Experts debate how quickly that will happen, because natural variability in the past 10–20 years may have accelerated the disappearance of ice caused by human-induced global warming. So ice loss could slow and ice cover might even rebound for brief periods. "At the end of the day, there's a lot of uncertainty," says Marika Holland, a sea-ice researcher at the National Center for Atmospheric Research in Boulder, Colorado.

But the long-term trend seems clear. "The consensus seems to be that out in the 2030s to 2050s or 2060s is when we might see the loss of all the ice for a short period during the summer," says John Walsh, an atmospheric scientist at the University of Illinois at Urbana-Champaign.

That is likely to lead to a significant rise in traffic in the Arctic Ocean. With the summer sea ice both shrinking in extent and thinning, the region is becoming much more accessible

to ice-strengthened ships, which have less formidable protection than true icebreakers. A study conducted by researchers at the University of California, Los Angeles, estimated that by mid-century 23% more of the Arctic's waters will be accessible to vessels capable of just limited icebreaking².

Eventually, shipping companies might start to use the Arctic as a short-cut for transporting goods between cities on the Pacific Rim and those bordering the Atlantic. Experimental voyages have been made along the north coast of Russia, and a smattering of ships has crossed the Northwest Passage north of Canada. But don't expect a significant rise in trans-Arctic traffic any time soon. In a 2009 assessment³, the Arctic Council, an intergovernmental forum for issues affecting the region, projected that most of the shipping in the region will involve bringing supplies to northern communities and exporting resources such as oil and minerals, for at least the next decade and possibly much longer.

"The notion that the Arctic Ocean will become a Panama Canal or a Suez Canal is a figment of the media," says Lawson Brigham, a geographer at the University of Alaska Fairbanks and chairman of the assessment. But, he adds, "there may be a short, summer 'window of opportunity' for trans-Arctic navigation".

These changes are creating a sense of urgency among scientists trying to answer a string of questions about the region (see 'Top questions in Arctic research'). Researchers seeking access to the Arctic Ocean have traditionally relied on icebreakers to get through the ice. But these ships are in short supply in the United States and, to a lesser extent, in Europe, because of lack of investment.

Even ice-strengthened vessels can be difficult for researchers to secure. "Because of the retreat of the sea ice and the oil development we have pending in the Chukchi and Beaufort Seas, a lot of the ice-strengthened vessels are being taken up by industry," says Jacqueline Grebmeier, an Arctic researcher at the University of Maryland in Solomons who has made several trips through the region on the Canadian Coast Guard icebreaker *Sir Wilfrid Laurier*.

Ice-strengthened vessels, meanwhile, are not sturdy enough to provide the kind of access that scientists most desire. "A lot of the processes that are really fundamental to the understanding of how Arctic climate, oceanography and biology work are not happening in summer time," says Lester Lembke-Jene, a marine geologist at the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, Germany. "If you don't have the full annual observation," he says, "you have a very, very imbalanced and narrow glimpse of what's going on there."

Given the shortage of vessels, researchers are trying to convince others to do some of the work on their behalf. Some commercial ship operators are collecting data for climate scientists, and last year the US government relaunched the Science Ice Exercise, in which military submarines



A researcher ventures off the coast of Alaska to measure the thickness of sea ice earlier this summer.

“THE NOTION THAT THE ARCTIC OCEAN WILL BECOME A PANAMA CANAL OR A SUEZ CANAL IS A FIGMENT OF THE MEDIA.”

NASA/K. HANSEN collect water samples and measure temperature, salinity, nutrients, chlorophyll and other data for scientists. Researchers are also starting to use autonomous underwater and aerial vehicles to collect data (see ‘New eyes for science’).

FROZEN OUT

The territorial ambitions of different nations may also end up restricting scientific access. In theory, the areas claimed under the UNCLOS apply only to the sea floor and do not give a country rights over the water above. In practice, however, such claims could hinder scientific work.

“If a coastal state wanted to, it could, by declaring regions to be of special interest for exploration, require that other states request permission to conduct research in the area of the extended continental shelf,” says Larry Mayer, a marine geologist at the University of New Hampshire in Durham.

This is more than just idle speculation, he says, because Russia has a history of impeding access to scientists from other nations seeking to work in its waters. Some researchers say that their attempts to put out or collect equipment from areas under Russian control have been thwarted when applications for permits were either denied or went unanswered. The Integrated Ocean Drilling Program, for example, could not obtain permission to drill in Russian parts of the Bering Sea in 2009.

Some scientists familiar with Russia say that the permission problems stem more from the nation’s massive bureaucracy than a deliberately obstructionist policy. “Inertia here coming from the Soviet era is really huge,” says Igor Polyakov, a Russian Arctic researcher who now works at the University of Alaska Fairbanks. He says that gaining permission for research in Russian waters is much easier now than in the past.

Others report just the opposite. Cheryl Rosa, deputy director of the US Arctic Research Commission in Anchorage, Alaska, says that researchers are still experiencing problems with permits, visas, taxations on funding, getting data out of Russia and other issues.

Despite these concerns with Russia, the view from scientists is that collaborations in the Arctic have never been stronger. The recent International Polar Year, which ran from March 2007 to March 2009, helped by bringing together scientists from different nations to work on dozens of projects in the Arctic, ranging from permafrost studies to research on indigenous peoples. “That’s really made a big difference and probably pushed science collaborations further and faster,” says Julie Brigham-Grette, a glacial geologist at the University of Massachusetts in Amherst.

Mayer hopes that the Arctic Council or some other body will forge an agreement that will allow researchers continued access across the Arctic. A positive step came in May when the council’s eight member states agreed to work together in search and rescue in the region, creating the first legally binding agreement made via the Arctic Council.

Arctic scientists say that the spirit of cooperation is changing their research in profound ways. They are probably ahead of their colleagues elsewhere in terms of releasing information quickly and widely, says Eicken. Data from the NSF’s Arctic Observing Network — a system of atmospheric and land- and ocean-based monitoring tools — for example, are made immediately available through an online database so that anyone can access near-real-time readings from the buoys and other observatories.

The advances are motivated in part by the changes in the region. Scientists often take two to three years to process and publish data, but in the Arctic, Eicken says, “that cycle isn’t able to keep up with the rapid development”.

And if projections are right, scientists will need to work faster than ever to keep up with the unstable conditions. Grebmeier saw just how changeable the region could be in July in her latest trip on the *Sir Wilfrid Laurier*. Several years ago, the ship had run into thick, multi-year ice when cruising near Barrow, Alaska. But this year, the crew found just one bit of ice and even that wasn’t terribly impressive. “I was working on the deck — someone yelled ‘sea ice,’” she says. “I looked up and actually thought taking a picture wasn’t worth my time.” ■

Daniel Cressey is a reporter with *Nature* in London. Additional reporting by Richard Monastersky.

1. Gautier, D. L. *et al. Science* **324**, 1175–1179 (2009).
2. Stephenson, S. R., Smith, L. C. & Agnew, J. A. *Nature Clim. Change* **1**, 156–160 (2011).
3. Arctic Council *Arctic Marine Shipping Assessment 2009 Report*; available at go.nature.com/jliv8x
4. Sohn, R. A. *et al. Nature* **453**, 1236–1238 (2008).

COMMENT



ARCTIC Scientists have much to learn from indigenous northern communities **p.182**

BIOETHICS Extreme politics unites biotechnology's enemies in America **p.184**

EXHIBITION How the skull has been revered since prehistory **p.186**

CLIMATE CHANGE A heated debate over sea ice minima **p.188**

H. MAERTENS FOTOGRAF



A green Arctic

Academic collaboration is essential for creating a sustainable future for Arctic development, says **Lars Kullerud**.

This summer, the University of the Arctic celebrated its tenth anniversary by asking whether the future of the north could be a green one.

This vision presents a challenge. Local communities want economic growth, but the easiest ways of achieving that goal are not necessarily sustainable. The north needs a route for development that isn't based solely on resource extraction. Academic collaborations can help to achieve this, by promoting knowledge-based development, and by answering the research questions needed to support sustainable development.

The Arctic is home to a vast wealth of resources. Covering about 7% of the globe (the United States, in comparison, covers 2%), the Arctic has a disproportionately large share of oil and gas (the US Geological Survey estimates that the Arctic holds 22% of the remaining undiscovered petroleum

resources), and includes large swathes of rocks rich in minerals. Diamonds and nickel are plentiful; the waters churn with fish; and the region is bordered by the vast boreal forest belt, which holds one-third of global forests and perhaps 40% of economic forest resources.

Arctic states are promoting the development of natural resources — sometimes with the active support of local people, at other times against local wishes — to secure national economic growth, stable access to resources and job creation. Many indigenous people support development, with provisos. In early 2011, the Inuit Circumpolar Conference — the umbrella organization for

Inuit peoples of the world — issued a declaration welcoming environmentally sound extraction of renewable and non-renewable resources as long as decisions are made locally and the economic benefit stays at home. A similar policy is already in effect in Greenland, actively promoting and welcoming oil and mineral exploration.

Unlike many of the world's conflict-ridden treasure troves — from the diamond-rich African nations to the Middle East's oil fields — the Arctic's resources are in the politically stable backyard of developed countries. But this does not eliminate concerns. Some non-Arctic states fear a future in which a sparsely populated north controls such a large portion of major resources that are in demand throughout the world.

Development initiatives are often met with public protests — from Greenpeace as well as some members of the European

ILLUSTRATION BY SOPHIE CASSON



AFTER THE ICE

Science at the new Arctic frontier
nature.com/arcticfrontier

► Parliament, non-Arctic nations and non-governmental organizations. Local governments and peoples in turn strongly object to attempts by outsiders to impose controls. The European Parliament's 2008 resolution on Arctic governance, which, inspired by the Antarctic Treaty, called for an international treaty to protect the Arctic, combined with the 2009 ban on seal products among other items, was so provocative, that the European Union's request to be an observer on the Arctic Council — the eight nations with territory in the Arctic — has repeatedly been put on hold. Many conservation organizations, such as Greenpeace and Sea Shepherd, and animal-welfare organizations, receive little support across the Arctic, whereas those that strongly encourage local dialogue, including the WWF, are more welcomed.

Academic institutions are working with governments, companies and other institutions to boost economic growth in the Arctic. Cooperation between the private sector, the University of Oulu and local government in Oulu, Finland, fostered the growth of the Finnish mobile-phone company Nokia. Likewise, universities in Tromsø, Norway, and Fairbanks, Alaska, are drivers of knowledge-based development; as are two new institutions in northern Russia: the Northern Arctic Federal University in Arkhangelsk and the North-Eastern Federal University in Yakutsk.

The University of the Arctic was started in 2001 as a network of collaborating higher-education institutions, supported by the member states of the Arctic Council and the Arctic indigenous peoples. Today it links 138 universities and colleges across Russia, North America and the Nordic countries, ranging from small institutions with 100 students to major research-intensive universities with tens of thousands of students. The network shares research initiatives and joint-study programmes, and helps to opti-

“Local peoples strongly object to attempts by outsiders to impose controls.”

mize use of limited resources. It is often argued that a critical mass of people and jobs in one location is necessary to create a viable economy in today's world. The

University of the Arctic proves that a dispersed network can be just as successful.

Another important role of academia is as the place where research questions are formulated and tackled. The International Arctic Science Committee, the University of the Arctic and the International Arctic Social Sciences Association have agreed to jointly organize the third International Conference on Arctic Research Planning in 2015. This is a bottom-up, scientist-driven initiative carried out every ten years to identify major research

questions important in the north. This time, we intend to include more contributions from local peoples. Their concerns are often different from those of academics: with climate change, for example, their focus is on building the knowledge necessary to deal with its effects, rather than on revealing more details of the processes behind it. Local peoples seek ways to merge traditional and academic knowledge to help develop insights that might, for example, be relevant to fishermen when fish migration patterns change, or to reindeer herders when grass and snow conditions alter, as well as replacing soot-producing fossil energy with renewable options.

It is only when the whole population of the Arctic has a say in developing knowledge in and about the region that we will be equipped with the tools to define our own future and decide at what pace our resources will be developed. It is crucial that academics, through facilities such as the University of the Arctic, become involved with those deliberations, so that the right scientific data are made available for policy-makers. Working together, we can strive towards a green economy in the north. ■

Lars Kullerud is president of the University of the Arctic, and is based at UNEP/GRID-Arendal, N-4836 Arendal, Norway.
e-mail: lars.kullerud@uarctic.org

A peaceful Arctic

Encourage dialogue between the producers and consumers of scientific knowledge in the north to keep the region conflict free, says **Oran R. Young**.

A cocktail of powerful forces, including the onset of climate change, the globalization of economic relationships and the shifting distribution of power in international society, is transforming the Arctic. Once regarded as a remote region of interest largely to explorers, missionaries and anthropologists, the Arctic has become a focus of attention for captains of industry and global policy-makers.

Conditions in the far north are very different from those at the opposite side of the globe. There, the highly effective Antarctic Treaty System relies on the scientific community to help administer the internationally agreed provisions for jurisdiction, demilitarization, environmental protection and the prohibition of mineral development in the Antarctic. There are no direct counterparts to this role in the Arctic, a region that is home to millions of human residents, subject to the undisputed sovereignty of its coastal

states, a theatre of operations for nuclear-powered icebreakers and naval vessels, and a site of world-class industrial activities including mining.

Journalists and pundits have broadcast dramatic scenarios that feature a scramble for the Arctic's resources, leading inexorably to resource wars and armed clashes. These concerns are misplaced. In reality, the eight Arctic states have settled most disputes over boundaries and the use of the region's resources through cooperative measures; they have also created the Arctic Council, a body that provides a forum for addressing emerging issues in an orderly manner.

Scientists have long played a part in these peaceful interactions in the Arctic, and they

will continue to do so. But steps can and should be taken to increase the relevance of science to emerging policy concerns, to improve the transfer of scientific knowledge and expertise into the hands of policy-makers, and to ensure that the Arctic remains a zone of peace.

BREAKING THE ICE

In 1987, Mikhail Gorbachev, then president of the Soviet Union, gave his 'Arctic zone of peace' speech, in which he called for a series of concrete measures to overcome East-West divisions in the area, including arms control measures and cooperative resource development. Gorbachev explicitly addressed the role of scientists in achieving that goal, and in the wake of his speech, science became an important vehicle for communication between the two camps. This led to the establishment in 1990 of the International Arctic Science Committee





(IASC) — a non-governmental group of national scientific organizations aimed at promoting cooperation.

Once the ice was broken, the Arctic states (Canada, Russia, the United States and the five Nordic countries) rapidly formed the Arctic Environmental Protection Strategy, in 1991, and its successor, the Arctic Council, in 1996. The most striking achievement of these bodies is the production and dissemination of a set of influential, science-rich reports, including *The State of the Arctic Environment* (1997), *Arctic Human Development Report* (2004), *Arctic Climate Impact Assessment* (2004) and *Arctic Marine Shipping Assessment*. The first report's evidence of pollutants in the pristine northern environment acted as a catalyst for the 2001 Stockholm Convention on Persistent Organic Pollutants.

Science continues to contribute to effective governance of the Arctic. First, it

provides an idiom for addressing contentious issues. Take, for example, the delimitation of Arctic coastal state jurisdiction over the seabed beyond the Exclusive Economic Zones. These states have rights to areas that are shown to be natural extensions of their continental shelves (see pages 172 and 174). Canada, Denmark and the United States have joined forces to conduct the mapping research needed to present evidence to the United Nations Commission on the Limits of the Continental Shelf. Russian policymakers have expressed a clear desire for scientific cooperation in resolving any jurisdictional disagreements.

Second, science helps to identify emerging issues and provide the evidence needed to move them towards the top of the policy agenda. A recent example involves the discovery of the prominent role of short-lived climate forcers, such as atmospheric soot ('black carbon'). The scientific

documentation of this problem prompted the Arctic Council to create a task force on the topic in 2009, leading to a 2011 Arctic Council decision to organize demonstration projects (for example on fire prevention) to help reduce emissions.

A third role for science is to provide a reality check for commercial ventures: researchers can amass evidence to assess the credibility of forecasts of oil and gas reserves in the Arctic, the possible obstacles to commercial shipping, and the requirements for managing commercial fisheries.

POLE TO POLE

Despite the differences between the north and south poles, northerners can learn from the Antarctic experience, notably the strong, institutionalized links between the producers of scientific knowledge and the guardians of the region's governance system.

The Scientific Committee on Antarctic Research (SCAR) is a non-governmental body operating as a committee of the International Council for Science (ICSU). It includes the national research councils and scientific academies of countries involved in Antarctic research, and its services are relied on by the Antarctic Treaty Consultative Meetings — the annual meetings of the governments of the treaty states. SCAR also collaborates with the Committee for Environmental Protection, the expert body that advises the politicians on environmental policy in the Antarctic. In the Arctic, the IASC has a comparable role to that of SCAR. But it is neither a committee of ICSU nor an explicitly acknowledged contributor to the work of the Arctic Council. This makes for a tenuous connection between the producers and the consumers of Arctic science. Communication between the two groups is complicated, and is hampering the development of mutual respect and trust.

Such problems almost torpedoed the production of a policy document summarizing the key findings of the Arctic Climate Impact Assessment for consideration at the 2004 Arctic Council ministerial meeting. The absence of a well-defined and mutually understood procedure benefited no one.

The IASC and the Arctic Council must, and will, remain separate, each with its own mandate, membership and operating procedures. But they need an explicit agreement on 'rules of engagement' that encourage communication. The next Arctic Council ministerial meeting in 2013 should adopt a memorandum of agreement covering this relationship. ■

Oran R. Young is at the Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, California 93106, USA. e-mail: oran.young@gmail.com



Sami herder Mattis Sara (right) helps Nicholas Tyler take a reindeer blood sample in Karasjok, Norway.

The local perspective

Indigenous knowledge is maturing as a science, says **Henry P. Huntington**. But more work is needed to give the field the respect it deserves.

In April 1995, I sat with a group of nine Iñupiaq and Yupik elders in the community centre in Koyuk, Alaska, documenting information that they had acquired — from experience, observation and previous generations — about beluga whales. At one point, the conversation took a surprising turn — from whales to beavers. I must have looked confused. One of the elders smiled and explained that a growing beaver population was damming streams in which fish spawn, affecting the belugas' food source¹. The connection was clear when spelled out, but it was not one that most whale researchers would have anticipated.

The concept of scientists using indigenous, or traditional, knowledge in their research has received increasing attention over the past few decades. This is particularly true in the Arctic, where the potential global

effects of changes such as permafrost thaw and ice melt have created an urgent need to understand how climate change is affecting the region. Historical physical data about the region are lacking, but indigenous cultures there have retained practices and knowledge acquired over countless generations.

More and more, scientists are finding value in collaborating with these populations. And growing political awareness and activism by indigenous peoples have led to increased recognition of their knowledge and ideas.

Yet there is still far to go. In 2007, for example, although the Intergovernmental Panel on Climate Change (IPCC) included discussions of traditional knowledge in its sections on Africa and Polar regions², it did not consider the topic for other parts of the world. And in these two cases, discussions focused on

societal adaptation to climate change rather than on sharing observational data.

In July 2011, a handful of organizations, including the United Nations University and the IPCC, convened a workshop in Mexico City in an effort to overcome the language and sociocultural barriers preventing inclusion of traditional knowledge in the research literature. But not all scientists are yet convinced of the value of this approach, nor do those who appreciate it always know how to pursue it.

SEA CHANGE

I first encountered traditional knowledge in the early 1990s while working for the Alaska Eskimo Whaling Commission in Barrow. Whenever I visited a village, I would open a map and ask the local whalers to introduce me to the area, so I would know the places they described when they called me later in the season to report their hunting activities. The maps always caused an excited stir among them as they showed me where the sea ice was in spring, where the whales swam and other details of local geography and ecology that were largely undocumented in the scientific literature. I was hooked.

Getting funding for a formal project to document traditional knowledge was a different story — in part because funding agencies were just beginning to recognize the legitimacy of the topic, and in part because I needed to learn how to frame such a project.

In 1994, after a failed proposal, I went to Anchorage, Alaska, to work for the Inuit Circumpolar Council (ICC), an organization that works to promote the value of indigenous knowledge to sympathetic-yet-sceptical scientists and managers. While there, I received funding from the US National Science Foundation. My research subject had switched from bowhead whales to beluga whales, but the basic idea remained the same.

A sea change was coming. Groups such as the ICC became more vocal, and scientists who spent time with indigenous peoples grew increasingly appreciative of what they had to say.

Today, there are many collaborations between scientists and Arctic residents. For example, Sami reindeer-herders in Sweden and Norway have worked with scientists to document snow conditions and their implications for herding practices, now and in future climatological scenarios³.

In Canada, climatologists and Inuit have uncovered subtle changes in weather patterns over the past few decades by



AFTER THE ICE

Science at the new Arctic frontier
nature.com/arcticfrontier

supporting qualitative observations with statistical correlations⁴ and identifying places to install weather instruments to capture locally relevant data. Sea-ice scientists and Inupiaq whaling captains have combined their perspectives to give a fuller view of ice patterns, dynamics and trends⁵. Marine mammalogists followed up on Inupiaq observations to discover that bowheads, unlike most other whales, have a sense of smell⁶.

The rhetoric among scientists and funding institutions has shifted from polite rejection to bemused tolerance and, increasingly, openness. In 2007, the North Pacific Research Board in Anchorage allocated US\$1 million to traditional-knowledge research as part of its \$14.7-million contribution to the 5-year, multi-agency Bering Sea Integrated Ecosystem Research Project.

And yet, there is much still to be done.

STRONG FOUNDATIONS

First, natural scientists must do more to engage with the topic. Social scientists, myself included, have done much to bring traditional knowledge to wider attention and develop appropriate methods for documentation and collaboration^{7,8}. But experts in other fields need to join the conversation.

In 2010, I attended a meeting in northern Quebec, Canada, at which Peter Kattuk, an Inuit hunter from Sanikiluaq in Hudson Bay, noted that the seals he had caught that winter had shrimps instead of fish in their stomachs, and tended to sink rather than float in the water. I felt the observation was important, but I lacked the knowledge to see why.

A few days later, I mentioned Kattuk's observations to Eddy Carmack, an oceanographer at Fisheries and Oceans Canada in Victoria, British Columbia. He recognized them as early warning signs of a food-web 'flip' in Hudson Bay, which might otherwise have taken years to detect. Once Kattuk and Carmack made contact, the real exchange began. We need to foster more of these interactions.

Scientists should, however, do some preparation before reaching out to local populations. Insensitivity to local concerns and misunderstandings about data ownership have caused projects to be cancelled or undermined. In the spring of 2002, for example, researchers had to make last-minute changes to the cruise plan for an icebreaker heading to northern Alaska because they had assumed (wrongly) that the lack of communication from locals about the voyage was a sign of consent.

Second, researchers should scrutinize traditional knowledge and its sources more carefully. Under the banner of political correctness, some scientists are slow to critique what is said by people from other cultures, which can undermine the

reliability of information and the credibility of the field.

In the late 1990s, for example, beluga whales in Cook Inlet in southern Alaska were in the midst of a sharp population decline. Some self-professed local experts (perhaps seeking attention, perhaps to avoid hunting restrictions) claimed that there were 50,000 whales in the Gulf of Alaska, but only a few hundred of them came into Cook Inlet at a time, creating a false impression of a crisis. Despite the lack of any corroborating information, few scientists, fishermen or recognized local experts were willing to challenge these claims publicly.

Indigenous cultures have ways of establishing credibility, such as being careful to include the source and format of any information they pass on. Researchers need to be similarly diligent.

Third, researchers and locals should collaborate to establish new ways of gathering traditional knowledge. In 2008, hunters in Nunavut, Canada, began carrying Global Positioning System units to record phenomena as they travelled their usual routes⁹.

The technique was so successful that the researchers and the Inuit now plan to travel in 2012 to Nepal, to share the method with



Inupiaq elder Warren Matumeak (left) and glaciologist Andy Mahoney examine sea-ice data.

residents of the Tsum Valley, which has lately seen increases in glacial retreat and other signs of climate change. The Inuvialuit in northwestern Canada are developing a monitoring programme for their region, addressing local interests as well as topics identified by visiting scientists. An important element of such projects is that participants are paid for their work, recognizing the importance of their contributions and providing for greater stability in personnel.

These forward-looking approaches add another dimension to the typically retrospective documentation of traditional knowledge, and should be supported by the same agencies that fund other scientific endeavours.

Finally, we need to develop better ways of managing this information. Videos, maps and songs are better-suited than spreadsheets for recording traditional knowledge, but those formats are difficult to search or make accessible. Researchers at the Exchange for Local Observations and Knowledge in the Arctic (www.eloka-arctic.org) in Boulder, Colorado, are among those who are developing solutions — using advances in multimedia and data-tagging schemes to better represent knowledge.

The question of ownership is another issue. In 1989, after the *Exxon Valdez* oil spill in Prince William Sound, Alaska, hunters and fishermen in the area had mixed responses to outsiders seeking their knowledge. Some saw benefits for all in collaborating; others expected a cut of future revenue in return for sharing their knowledge; still others worried about sensitive cultural information becoming public.

As a result, many research projects were delayed and proposals abandoned as protocols were debated. In 2010, in contrast, Yupik elders on Nelson Island in southwestern Alaska made public their database of local land-use patterns and environmental conditions, on the principle that sharing was of paramount value.

Data-sharing practices should acknowledge cultural sensitivities and work in both directions. Researchers should feed information back to hunters, elders and other locals, so that they can join the conversation and see for themselves how their observations connect with other data. At its heart, the greater engagement of scientists with traditional knowledge is really about the greater engagement of people, sharing information about what matters most to them. ■

Henry P. Huntington is Arctic science director for the Pew Environment Group, 23834 The Clearing Drive, Eagle River, Alaska 99577, USA.
e-mail: hhuntington@pewtrusts.org

- Huntington, H. P. & the Communities of Buckland, Elim, Koyuk, Point Lay and Shaktoolik *Arctic* **52**, 49–61 (1999).
- Parry, M. L. et al. (eds) *Climate Change 2007: Impacts, Adaptation and Vulnerability* (Cambridge Univ. Press, 2007); available at go.nature.com/dgddcm
- Riseth, J. Å. et al. *Polar Rec.* **47**, 202–217 (2011).
- Weatherhead, E., Gearheard, S. & Barry, R. G. *Global Environ. Change* **20**, 523–528 (2010).
- Mahoney, A., Gearheard, S., Oshima, T. & Qillaq, T. *Bull. Am. Meteorol. Soc.* **90**, 370–377 (2009).
- Thewissen, J. G. M., George, J., Rosa, C. & Kishida, T. *Mar. Mammal Sci.* **27**, 282–294 (2011).
- Briggs, C. L. *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research* (Cambridge Univ. Press, 1986).
- Huntington, H. P. *Arctic* **51**, 237–242 (1998).
- Gearheard, S., Aipellee, G. & O'Keefe, K. in *SIKU: Knowing Our Ice: Documenting Inuit Sea-Ice Knowledge and Use* (eds Krupnik, I., Aporta, C., Gearheard, S., Laidler, G. J. & Kielsen Holm, L.) 181–202 (Springer, 2010).



J. REED/REUTERS

One of the most heated debates in US politics is over the use of human embryonic stem cells. Here, George W. Bush in 2006 holds a child born from a frozen embryo.

BIOETHICS

Brave new biopolitics

Kevin Finneran hails a timely take on the debate raging over biotechnology breakthroughs in the United States.

From stem cells to synthetic organisms, advances in the life sciences are as likely to set off a conflagration of debate as a celebration of progress. Such “new biology”, says Jonathan Moreno in *The Body Politic*, challenges our self-perceptions, social values and even political systems. The age of bioscience has become the age of biopolitics.

Moreno, a bioethicist at the University of Pennsylvania in Philadelphia and a fellow at the Center for American Progress in Washington DC, explores this political battlefield in depth. He examines today’s wars, and also the historical and philosophical streams that have fed opposing thought about biotechnologies in the United States. His aim is to develop a richer “moral conversation” about bioscience. Moreno devotes much of the book to a critique of what he sees as a neoconservative hostility to science, and explains how science can be a key ingredient of a progressive political agenda.

He charts the issues well, providing clear summaries of the use in research of

pluripotent stem cells derived from human embryos, the implications of the successful cloning of Dolly the sheep and the possible applications of synthetic biology. He covers the use of life-support technology to sustain the lives of patients with severe brain damage; the blurry line between therapeutic and enhancing interventions in human reproduction; and the disconcerting history of eugenics.

Moreno shows how developments in biotechnology have affected people across the ideological spectrum. Over the past decade, these have led to new alliances between political factions and a difficulty



**The Body Politic:
The Battle
Over Science in
America**

JONATHAN D.
MORENO

Bellevue Literary Press:
2011. 224 pp.
\$18.95, £12.99

in finding common ground between others. The unlikely confraternity that questions the directions taken by life scientists includes conservatives concerned about abortion, neoconservatives worried about threats to human dignity and liberals fretting that new biotechnologies will exacerbate existing economic inequality. In this alliance, people on both the left and right sides of the political divide claim to be motivated by a preference for what is ‘natural’. For the right, this usually means conventional human reproduction and the protection of human dignity. For the left, it is species and ecosystems.

Moreno does not engage with arguments from religious conservatives; faith in divine guidance is difficult to refute. His fight is with the neoconservatives, especially those linked with the administration of former US president George W. Bush. Moreno criticizes what he sees as the

NATURE.COM
Harold Varmus on a
guide to US science
and government:
go.nature.com/ad1jfa

neoconservatives' fundamental distrust of human nature, which underlies their belief that people cannot manage technological progress wisely. He makes much of the irony that neoconservatives associate science and technology with human commodification and alienation — the very problems once identified as the hallmarks of capitalism by Karl Marx. Furthermore, he shows that the neoconservative analysis is too shallow: progressives go farther by concluding that the social and economic backdrop to a technology is what determines its impact.

Moreno goes to some lengths to place his ideas in a larger philosophical context that extends from Socrates to Friedrich Nietzsche. In particular, he aims to demonstrate that a progressive, pro-science stance is consistent with the values of the eighteenth-century Enlightenment period that inspired the birth of the United States and the development of other Western democracies.

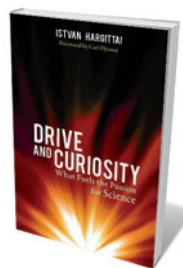
But Moreno's analysis focuses too heavily on the neoconservatives. He does not address why US Republicans do not object to the use of biotechnology in agriculture or pharmaceuticals, to nuclear power or to new computer and communications technologies. The 'neocon' label in US politics has changed in meaning since it was first coined, and it is now associated mainly with an interventionist foreign policy. Are the main figures from the Bush era — such as Leon Kass, who headed the President's Council on Bioethics from 2001 to 2005 — still influential? How many conservatives would applaud Kass's statement in his book *Toward a More Natural Science* (Free Press, 1985): "Science — however much it contributes to health, wealth and safety — is neither in spirit nor in manner friendly to the ... moral and civic education of human beings and citizens"?

Moreno frets that the political right wing, by investing in the human factor — such as reproduction — has the advantage in appealing to the public. He encourages the left to build on its concern for human rights and the redistribution of wealth, and to create a progressive biopolitics.

The United States itself provides a model for how allowing scientists to explore and test new ideas can result in benefits for all, says Moreno. The country is an enormous social experiment that evolves by learning from the evidence of what works, and operates through fundamental values such as openness to new ideas. The challenge is to maintain this human side of science when the research, to many people, seems to be a threat to what is essentially human. ■

Kevin Finneran is editor-in-chief of *Issues in Science and Technology*, a publication of the US National Academies and the University of Texas at Dallas.
e-mail: kfinnera@nas.edu

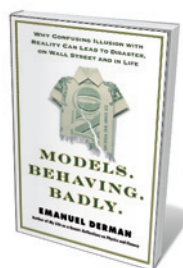
Books in brief



Drive and Curiosity: What Fuels the Passion for Science

Istvan Hargittai PROMETHEUS 338 pp. \$26 (2011)

What propels scientists towards great breakthroughs? Drive and curiosity are only part of it, argues Hungarian chemist Istvan Hargittai. Personality, motivation and accidents of circumstance led luminaries such as the co-discoverer of DNA structure James Watson, Nobel-prizewinning biochemist Gertrude Elion and theoretical physicist George Gamow to their big discoveries. The snapshot of Frederick Sanger — two-times Nobel prizewinner — is refreshing. The methodical biochemist's regard for doing a steady job, says Hargittai, helped to pave the way to the Human Genome Project.



Models. Behaving. Badly: Why Confusing Illusion with Reality Can Lead to Disaster, on Wall Street and in Life

Emanuel Derman FREE PRESS 240 pp. \$26 (2011)

As a theoretical physicist and Wall Street analyst, Emanuel Derman has produced financial models that have become industry standards. But he has long warned that such models, however elegant, are inadequate: human behaviour must be factored in. Derman calls for financial models to be seen as "parallel thought universes" that should be taken with a kilogram of salt. Ranging wittily across philosophy, literature and the arcane world of high finance, Derman's argument is a heady mix of physics, economics and memoir.



Survival of the Beautiful: Art, Science, and Evolution

David Rothenberg BLOOMSBURY 256 pp. £14.99 (2011)

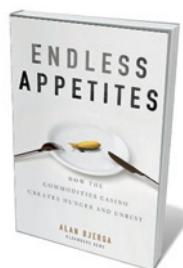
The colour blue rules for the male satin bowerbird of Australia. The interior decorators of the avian world, they gather plastic, shells and feathers of that hue to adorn their meticulously built stick structures, all to lure a potential mate. This is just one indication, argues philosopher and musician David Rothenberg, that beauty is not random but is intrinsic to life — and that evolution proceeds by sumptuousness, not by utility alone. Rothenberg covers topics such as camouflage, abstraction, the profound impact of art on science and much more to explore his theme.



1Q84: Book One and Book Two/1Q84: Book Three

Haruki Murakami HARVILL SECKER 624 pp./368 pp. £20/£14.99 (2011)

This Japanese sci-fi blockbuster, now translated into English by Jay Rubin and Philip Gabriel in two volumes, begins irresistibly. Aomame, a clinically precise assassin, escapes a traffic jam by climbing down an emergency exit — and, like Alice, enters a parallel universe. Meanwhile, mathematics teacher Tengo is rewriting a novel about a two-mooned world ruled by a tribe of 'little people'. Over one year, the lives of Aomame and Tengo slowly converge. Like its near-namesake 1984, it is a dystopian tale of boy meets girl. It is also a fevered, enigmatic journey into alternative realities and near-past Japan.



Endless Appetites: How the Commodities Casino Creates Hunger and Unrest

Alan Bjerga JOHN WILEY 208 pp. £18.99 (2011)

Globalization has spawned a crop-market 'casino' in which rich countries gamble over crop prices while millions go hungry. If it can be overcome, says Alan Bjerga, farmers could feed, clothe and fuel the world. Bjerga meshes number-crunching with research gleaned from trading floors in Chicago and farmers in Nicaragua, Thailand and Kenya. Feeding these countries' hardscrabble hectares into the global market is, he concludes, an achievable way to solve hunger.

ANTHROPOLOGY

Centuries of skulls

Alison Abbott enjoys a German exhibition charting how the human head is revered by cultures worldwide.

Skulls hold an abiding fascination for the living. From goth rings and Mexican Day of the Dead paraphernalia to artist Damien

Hirst's diamond-studded platinum cast, skulls are stark and emptied reminders of mortality — and repositories for meaning.

A German exhibition now reveals that skull cults have existed in nearly every culture since prehistoric times. *Schädelkult (Skull Cult)*, at the Museum of World Cultures in Mannheim, one of the Reiss Engelhorn Museums (REM), shocks and delights with its tales of how tribes around the world and across time prepared and used the skulls and heads of honoured ancestors or feared enemies.

The exhibition hints at why the skull so grips our imagination. It is the scaffolding for faces of those once loved, feared or respected — and a casing for their psyche, memories and social power. It is the starkest proof that an individual is no longer here. Every culture tries to rationalize death; many have tried to beat it. The skull has had different symbolic roles in these endeavours.

Schädelkult (Skull Cult)

REM Museum of World Cultures, Mannheim, Germany. Until 29 April 2012.

The oldest object, found in 1997 near Koblenz in northern Germany, is a Neanderthal skullcap — apparently used as a bowl — with an estimated age of 170,000 years. But even stranger exhibits abound. Skulls of respected tribe members might be made into jewellery to be worn by their descendants; decorated with carving, feathers or shells and brought out in rain-making rituals; or modelled in clay to create a true or idealized portrait.

Some African tribes, believing that the power of their slain enemies lived on in their skulls, made musical instruments from them for use in battle or to weaken hostile neighbours. They detached the lower jaws to make trumpets and used the craniums as drums.

Some of the most famous trophy heads were not skulls at all. The Jivaro people of the Andes developed a unique way of processing the decapitated heads of their enemies: after carefully skinning them, they would fill the skins with hot sand to shrink them, closing the eyes and mouth with bamboo needles to stop the soul from escaping. Wearing a shrunken head was thought to bring all-round success — in health, fertility and hunting. The Maori of New Zealand mummified their ancestors' heads — the ornately tattooed head of a chieftain, brought to Europe by James Cook in 1770, is among the finest exhibits.

The practice of artificially deforming skulls was surprisingly widespread, from central Asia to central Europe. Startling examples include those created by binding the heads of young children in bandages or bonnets to force them to grow into a spherical shape, or by squashing

the heads between planks to force them into more flattened 'table' or 'tower' forms.

Europe had its own skull cult. During the Middle Ages, ancestral skulls in parts of Alpine Europe were sometimes painted with floral patterns — a token of respect for bodies that had been disinterred to make

room in cemeteries for the newly dead. Skulls appear constantly in post-Renaissance Christian art and architecture as

a reminder that death can pounce unexpectedly, so it was wise not to jeopardize a

place in heaven by deviating from the path of

righteousness. A carved bust of the

Virgin Mary, for example, doubles up as a skull reliquary. *Schädelkult* acts as a

reminder that Western anthropologists have been collecting, measuring and interpreting skulls for some 200 years. The extensive ethnological collection of German artist and Darwinist Gabriel

von Max (1840–1915) spawned this exhibition

when several hundred of its skulls, thought to have been destroyed in the Second World War, were rediscovered in 2008 and returned to the REM.

Von Max had intended his collection to celebrate the diversity of humanity. But many collections were exploited in the nineteenth and early twentieth centuries for dubious phrenological and anthropometric studies. These fed erroneous theories claiming, for example, that particular races were inferior. The exhibition confronts this unscientific research, along with the sensitive question of whether present-day requests for skull repatriation should be heeded. The curators argue that this should only happen when living relatives request skulls, or if they are proved to have been acquired through murder.

The REM Museum of World Cultures is well equipped for forensic analysis and, together with collaborators, has carried out tests on some of the exhibited skulls to date and characterize them. Some of the results, including facial reconstructions, are displayed on screens around the exhibition.

The last showcase in the exhibition is packed with contemporary skull-bedecked objects — from furniture to a saucy mini-dirndl costume. Skull cults today ooze an irony that they never had in the past. ■

Alison Abbott is Senior European Correspondent at Nature.



A Peruvian trophy head from AD 100–600.



A shell-decorated ancestral skull of the Asmat people of New Guinea.

H. MAERTENS

J. CHRISTEN/REISS-ENGELHORN-MUSEUMS



Q&A Joichi Ito

Internet visionary

As the new director of the Media Lab at the Massachusetts Institute of Technology (MIT) in Cambridge, Joichi Ito brings his knowledge of Internet start-ups — including Flickr, Twitter and licence provider Creative Commons — to the lab that developed the ideas behind the game Guitar Hero and Amazon Kindle's E-Ink technology. Ito talks about the value of playfulness and freedom in scientific discovery.

How did you get interested in technology?

As a child, I wanted to become a physicist but spent more time playing with computers. My father was a Japanese polymer chemist recruited by a Michigan company, Energy Conversion Devices. As a teenager, I wrote software to measure responses to antireflective coating on photovoltaic cells for the firm. My mum was a college dropout and housewife, then joined the company, took us to Japan, and became president of its Japanese subsidiary. Later I was on the board, but I gravitated towards computers.

You dropped out of college twice. Why?

In the mid-1980s you could learn more about computers on your own than in the classroom. I dropped out of a computer-science

programme at Tufts University in Boston, Massachusetts, in 1985, and out of a physics course at the University of Chicago, Illinois, in 1987. When I started working as a DJ in nightclubs, digging into the mess of AIDS-ridden working-class Chicago, university life seemed rather shallow. As a high-school student I had a great mentor, the physicist and inventor Stanford Ovshinsky, and it was hard to find that kind of stimulation at university. If I'd had Richard Feynman as a physics professor, I would probably have been fine.

How will you convince MIT students to finish their PhDs?

Ultimately, I want to admit and retain people like me, who would otherwise drop out because they want to build things and

pursue projects. I jokingly asked MIT administrators about enrolling myself, but they said it would be awkward for the director to also be a student. Universities provide incentives to learn, but they have been designed on a scarcity model, where you pick only the smartest people and educate them, for a price. With the Internet, education should be more abundant and universities can become more open.

What are your research priorities?

I have my own biases but I'm not going to pursue my own research. The Media Lab itself is my work; everyone's work is my work. Hopefully we'll pursue even more diversity, including reaching out to hard-science departments such as physics and chemistry. We could collaborate more with theoretical scientists, such as anthropologists and sociologists.

How do you plan to change the Media Lab?

Media Lab scientists could get out and collaborate even more than they currently do. And people should come in, so the Lab is more of a platform than a container. Our corporate sponsors tend to be big firms and consumer-products companies, but we could also invite more non-profit organizations, foundations and others to collaborate.

Is there a downside to corporate-sponsored research?

Many companies pay consultants millions of dollars to massage their brains to make them feel a little smarter. The amount of money they invest at MIT is small by comparison. But transactional relationships can be difficult. You have to pick sponsors who want to be inspired by you, not just use you for your intellectual property. Hopefully there are enough of those to keep us afloat.

Do you share the Media Lab's values of playfulness and freedom?

I named my venture-capital firm Neoteny (<http://neoteny.com>), which in developmental biology means the retention of childlike attributes in adulthood. As adults, we tend to stop learning and focus on self-protection and retention, but as a child it is easy to feel wonder, to learn and adapt. Schools often teach children that they are stupid. Some teachers can cut children off, limiting their creativity and promoting self-censorship. The Media Lab attracts people who have not limited themselves, and encourages them to continue not to. In a changing world, activities that sound whimsical can be a great way to invest in your own ability to adapt. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Sea-ice minimum is not a one-off

There is heated discussion over whether or not this year's seasonal sea-ice minimum in the Arctic constitutes a record (see go.nature.com/bmydx7 and, for example, go.nature.com/286mdp). But the debate misses the crucial point: this year's minimum is evidence that the unprecedented seasonal minimum of 2007 was not a one-off.

Continuous global satellite sea-ice observations began in 1972. A value for sea-ice coverage is retrieved from the surface microwave emission, using the different emission properties of water and ice. The sea-ice extent is conventionally defined as the sum of all data elements that have more than 15% ice cover.

Several remote-sensing groups work on this topic, using various methods and data from at least two different satellite instruments: the Advanced Microwave Scanning Radiometer — Earth Observing System (AMSR-E) on NASA's Aqua satellite and the Special Sensor Microwave Imager/Sounder (SSMIS).

All of these groups agree that the seasonal 2011 minimum is very close to the 2007 minimum. That value was about 25% less than the previous low in 2005, and almost 40% less than the climatological mean for the seasonal minimum (1979–2007). In autumn 2007, this value could be considered an outlier, caused by unusually warm weather over large parts of the Arctic Ocean.

However, the five seasonal minima since 2007 are the lowest on record. Although it is too early to speak of a trend, other observations, such as the thinning of Arctic sea ice over the past two decades, also suggest that the 2007 and 2011 minima are not single outliers. **Christian Melsheimer, Georg Heygster, Justus Notholt**



*Institute of Environmental Physics, University of Bremen, Germany.
heygster@uni-bremen.de*

Vaccine research loses out

The US government has spent an estimated \$60 billion on biodefence since the terrorist attacks of 11 September 2001 (*Nature* 477, 150–152; 2011). This is an unprecedented investment in biomedical research and infrastructure.

The value of capacity-building and spin-off discoveries that advance drug discovery and vaccine research should not be underestimated. However, suppose a fraction of this money had been used to accelerate vaccine research for existing diseases that pose a clear and present danger. In the case of HIV, tuberculosis and malaria — which together kill around 6 million people every year — we might have vaccines for all three by now.

Of the billions that are spent each year on combating infectious diseases using the tools we have, less than 2% is

invested in vaccine research. The science is there to develop vaccines against these killer diseases, but the resources are not. A more balanced approach to allocating resources is needed.

Peter Hale *The Foundation for Vaccine Research, Washington DC, USA.*

peter.hale@vaccinefoundation.org
Simon Wain-Hobson *Institut Pasteur, Paris, France.*

Robin A. Weiss *University College London, UK.*

The JASONs bought into the Vietnam War

I was surprised that Ann Finkbeiner did not mention the JASONs' involvement in the war in Vietnam in the 1960s, in which they were far from independent (*Nature* 477, 397–399; 2011).

The JASONs, a secretive group of US defence science advisers, bought the Pentagon line that the problem was infiltration from North Vietnam. They proposed laying a carpet of anti-personnel weapons across the entire demilitarized zone, including gravel and spider mines, button

bomblets, and nail and cluster bombs. The US military replaced the 2.5-centimetre metal flechettes in the nail bombs with sharp slivers of plastic that Vietnamese physicians could not locate with X-rays.

As one critic said shortly before the war ended (D. Shapley *Science* 179, 459–505; 1973), the JASONs were “totally unaware that they're just being used by some little, puddin'-headed guy in the Pentagon”.

Joseph Schwartz *London, UK.*
josephschwartz@btinternet.com

Fallacy in costs of pollution control

You perpetuate a well-worn economic fallacy in arguing that tightening pollution regulations would help the economy because “money spent on reducing emissions does not disappear into a vacuum: pollution control is a business, too” (*Nature* 477, 249; 2011). As economist Frederic Bastiat pointed out in 1850, “The error begins when the sacrifice itself is said to be an advantage because it profits somebody” (*Essays on Political Economy* Provost, 1874).

If it costs money to reduce pollution, that is because we must pay to divert people and equipment away from other activities that also benefit society. The benefits of pollution control may be worth the costs, but it is pure doublethink to claim that costs are benefits.

Neal Hockley *Bangor University, Wales, UK.*

neal.hockley@univ.bangor.ac.uk

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the author guidelines at go.nature.com/cmchno. Readers are also welcome to comment online: www.nature.com/nature.

Fruity aphrodisiacs

Some fruit odours sexually arouse male fruitflies. The response is mediated by olfactory neurons that are sensitive to food smells and plug into the brain's neural circuit for sexual behaviour. [SEE LETTER P.236](#)

BENJAMIN PRUD'HOMME
& NICOLAS GOMPEL

The smell of a delicious stew often stimulates a man's appetite, but it rarely turns him on. Male *Drosophila* fruitflies, however, behave differently. On page 236 of this issue, Grosjean *et al.*¹ identify how particular odours from rotten fruit and decaying vegetables — the creatures' typical diet — act as aphrodisiacs to male fruitflies.

Males and females of many animal species produce chemical compounds that they use to find each other and to communicate their mating status. These volatile pheromones diffuse, and putative mating partners can detect them over fairly long distances. Female moths, for instance, release sex pheromones that can attract males from several kilometres away. How the nervous system perceives and interprets pheromones, and how these chemicals modify behaviours, are central questions in neurobiology.

Drosophila melanogaster is a powerful genetic model for addressing such questions. A key aspect of these flies' ecology is that they meet and mate where they feed. A series of cues from the female, including a pheromone bouquet spread on her abdomen, elicits sexual arousal in the male fly².

Grosjean *et al.*¹ find that odours unrelated to the female also promote male courtship behaviour. They show that phenylacetic acid and phenylacetaldehyde — two aromatic compounds found in fruits and other plants — specifically activate an olfactory receptor, IR84a, in sensory neurons on the fly nose (the antennae), thereby triggering male courtship behaviour. Indeed, when the authors removed the *Ir84a* gene, the mutant male flies not only couldn't smell these two odours, but also did not engage in courtship as vigorously as normal flies. Conversely, perfuming an unattractive mating partner (a dead female) with phenylacetic acid enhanced sexual appetite in the normal male, but not in the IR84a mutants.

How can sensory neurons tuned to detect

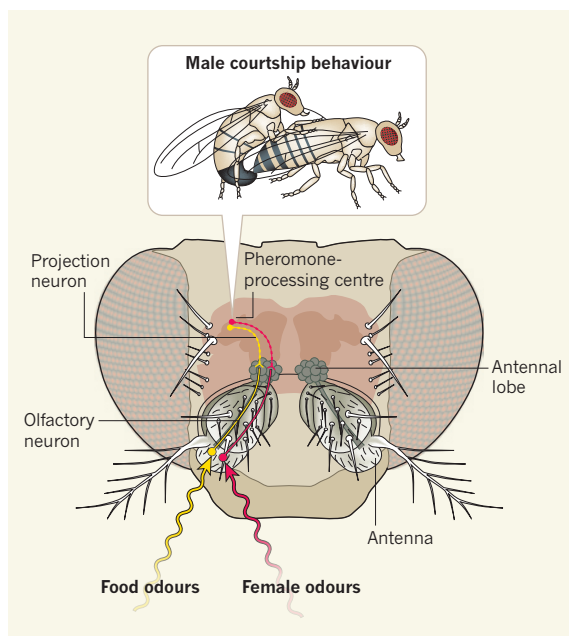


Figure 1 | When neurons meet. The male fruitfly's antennae perceive fruit-derived odours, as well as female odours. Dedicated sensory neurons convey this information to the antennal lobe, from where it is conveyed by projection neurons to a higher brain region — the pheromone-processing centre. There, the information from the two sensory stimuli intermingles, leading to male courtship behaviour.

food-derived odours promote sexual behaviour? Olfactory neurons connect to a brain centre called the antennal lobe, from where their projections are further parsed into smaller centres called glomeruli³. Grosjean and colleagues find that all IR84a-expressing olfactory neurons innervate the same glomerulus — one of the three glomeruli that are involved in male courtship behaviour.

Male fruitflies have a dedicated and complex neural circuit that controls most aspects of their courtship behaviour — from sensory perception to motor output^{4,5}. The identity of this courtship circuit is governed largely by male-specific isoforms of the transcription factor Fruitless (*FRU^M*). As it turns out, the IR84a olfactory neurons also express *FRU^M*. What's more, they form synaptic connections in the antennal lobe with projection neurons that also express *FRU^M*.

The projection neurons downstream of the

olfactory neurons dispatch different forms of sensory information to distinct brain regions. For example, information on food odours typically reaches one place and pheromone information goes to another. Yet Grosjean *et al.* show that the projection neurons downstream of IR84a sensory neurons are an exception: their axonal processes intermingle tightly with other *FRU^M*-positive projection neurons in the pheromone-processing centre of the brain (Fig. 1). In other words, the neural circuit mediating male-fly courtship is equipped with a plug-in to sense food resources that modulates the decision to court females.

How various sensory signals coming from different sources are integrated and computed in the brain into an actual decision is poorly understood. It seems, however, that the projection neurons conveying food or female-related odours connect to a unique centre that determines whether the olfactory environment is amenable to courtship. Knowledge of how this neural circuit reaches 'decisions' and how it triggers behaviours will require detailed characterization of the wiring and activity of the circuit^{6,7}.

Regardless of how male flies integrate food odours into the decision to mate, one may wonder why they do so in the first place. Female flies do worry about where they lay their eggs, but their male counterparts usually seem to have little say in this. Or perhaps they do. Linking courtship behaviour to a food compound gives the male some sway over how his progeny will be raised. Should the levels of phenylacetic acid be limited (read: low food resources), a male might not insist on courting local females, no matter how attractive they smell, and might wait to sire progeny grown in a richer environment.

This solution — linking food odour to male reproductive behaviour — has probably evolved relatively recently. Although IR84a belongs to an ancient family of receptors, it is found only in fruitflies, and is notably absent from insect species that rely on long-range



Figure 2 | Romantic setting. Male fruitflies not only feed on rotten fruit and vegetables, but also look for mating partners there.

pheromones to attract mates. Thus, evolution of IR84a may have shaped the ecology of fruitflies: targeting good food may have become a proxy for finding a sexual partner, thus circumventing the need for long-range pheromones (Fig. 2).

Advertisers often exaggerate the powerful role of human pheromones and perfumes in sexual attraction. It is nonetheless remarkable that the very chemical compounds that, according to this paper¹, enhance fly sexual appetite are also pleasant to humans and are used in perfumes for their sweet, floral smell. It remains to be seen to what extent environmental stimuli affect sexual behaviours in other species, including humans. ■

Benjamin Prud'homme and Nicolas Gompel are at Aix-Marseille University, Institut de Biologie du Développement de Marseille-Luminy, CNRS UMR 6216, Case 907, Parc Scientifique de Luminy, 13288 Marseille Cedex 9, France. e-mails: benjamin.prudhomme@univmed.fr; nicolas.gompel@univmed.fr

1. Grosjean, Y. *et al. Nature* **478**, 236–240 (2011).
2. Billeter, J.-C., Atallah, J., Krupp, J. J., Millar, J. G. & Levine, J. D. *Nature* **461**, 987–991 (2009).
3. Marin, E. C. *et al. Cell* **109**, 243–255 (2002).
4. Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirián, L. & Dickson, B. J. *Cell* **121**, 795–807 (2005).
5. Demir, E. & Dickson, B. J. *Cell* **121**, 785–794 (2005).
6. Ruta, V. *et al. Nature* **468**, 686–690 (2010).
7. von Philipsborn, A. C. *et al. Neuron* **69**, 509–522 (2011).

APPLIED PHYSICS

Controlled incandescence

Sources of incandescence emit their radiation across a broad spectrum. The finding that metamaterials can be used to tune the sources' emission spectrum makes these emitters an attractive prospect for some applications.

JEAN-JACQUES GREFFET

When a material is at high temperature, it emits light — a phenomenon called incandescence. This is how tungsten filaments in light bulbs produce light. Yet, for most applications, these thermal sources of light are inefficient because they radiate over a broad spectrum. Writing in *Physical Review Letters*, Liu *et al.*¹ describe how metamaterials can be used to design a thermal source that has a controlled, narrow emission spectrum in the infrared regime. Metamaterials are materials that have been engineered to contain nanostructures

that give them novel properties.

As Max Planck demonstrated in 1900, at thermal equilibrium, the radiation spectrum is independent of the material from which an object is made or its geometry, and depends only on temperature. However, in a non-equilibrium situation such as that of a light bulb, the spectrum is modified by the properties of the source's material. Progress in making incandescent light sources has thus been limited to choosing the best available material. So technologies based on phenomena other than incandescence are used to produce more efficient sources than thermal emitters, such as light-emitting diodes or fluorescent tubes

for lighting applications. Liu and colleagues' finding — that by using metamaterials it is possible to engineer the emission spectrum of an incandescent source in the infrared — paves the way for new applications for thermal sources. The authors describe sources that emit light at one or two wavelengths, or over a controlled spectral band, thus demonstrating good control of the emission spectrum. So how does it work?

The key to understanding the design of a thermal source is to apply Kirchhoff's law. This states that the thermal emission produced by a device is proportional to its absorptivity. Although it was introduced historically as a principle, the law has been mathematically derived^{2,3}, and so we now know that it is valid for any absorbing structure. Hence, designing an efficient emitter amounts to designing an efficient absorber, or in other words, to making light enter an absorbing structure without being reflected. With the development of nanophotonics, there has been a blossoming of new ideas for thermal sources, such as materials with periodic structure^{4,5} combined with surface resonances⁶, multi-layer systems⁷, metalodielectric systems⁸ and nanowires⁹. Such structures allow the absorption spectrum and the direction of emission to be controlled.

A standard technique for obtaining large absorption/emission is to use a thin metallic film separated from a mirror by a thin transparent layer. It was first introduced to provide anti-reflection coatings at radar frequencies (the 'Salisbury screen' concept) and is often used in infrared detectors. By replacing the thin film with small disks, quasi-monochromatic absorbers/emitters have been obtained¹⁰. Liu *et al.*¹ now introduce a metamaterial that consists of a periodic array of metallic crosses separated from a metallic mirror by a transparent, thin, insulating layer. The crosses are useful because they allow the emission frequency to be tuned by changing the length of their arms, just as the pitch of an organ pipe depends on the pipe's length. Interestingly, the authors show that it is possible to generate two absorption frequencies by combining two types of cross, each type having a different arm length (Fig. 1). They demonstrate an almost 100% absorption/emission for the two frequencies.

This large absorption might at first seem surprising, because crosses absorbing at one frequency cover only half of the sample's surface and so would be expected to absorb only half of the incident radiation. This apparent paradox is easily resolved by recognizing that, at subwavelength scales, light cannot be described in terms of optical rays propagating along straight lines. A cross illuminated at its peak absorption frequency (resonant absorption) can absorb as much radiation as an opaque disk that has an area (the 'absorption cross-section') much larger than the area of the cross. For example, the absorption cross-section of

an atom at the resonant wavelength λ takes the value $3\lambda^2/2\pi$, which is much larger than the size of the atom. Hence, it is possible to pack together several subwavelength-sized crosses that have different absorption frequencies but do not cover the entire sample's surface and still attain almost total absorption for all of these frequencies. This property is the key to the design of broad-spectrum absorbers/emitters using metamaterials.

What applications might there be for incandescent sources with controlled infrared spectra? Infrared thermal sources are usually used for identifying gases. Most gases have specific infrared absorption lines, so producing infrared sources designed to emit at these frequencies is an attractive idea. Another promising application is the development of thermophotovoltaic solar cells, which convert solar radiation into electricity. One of the main limitations to the efficiency of photovoltaic solar cells is the mismatch between the broad emission spectrum of the Sun and the narrow absorption spectrum of the photovoltaic cell. The thermophotovoltaic concept is based on having a broadband absorber that captures the Sun's energy and heats a secondary source that has an emission peak perfectly matched with the absorption spectrum of the solar cell.

Finally, Liu and colleagues' study¹ raises the prospect of active control of the thermal emission. The only active control of emission achieved so far has been based on modifying the temperature of the source. This is usually a slow process because temperature changes are limited by the time needed for heat to diffuse through the system. But if the resonant

absorptivity can be controlled by electrically or optically tuning a resonant absorber, a new generation of smart emitters can be envisaged. Liu *et al.* suggest that it would be possible to do just that to change the emission frequency. Another possibility is modulating the absorptivity/emissivity without modifying the source temperature. This could allow fast modulation of emission that is not limited by temperature dynamics. ■

Jean-Jacques Greffet is at the Laboratoire Charles Fabry, Institut d'Optique, CNRS, Université Paris Sud, 91127 Palaiseau Cedex, France.
e-mail: jean-jacques.greffet@institutoptique.fr

1. Liu, X. *et al.* *Phys. Rev. Lett.* **107**, 045901 (2011).
2. Rytov, S. M., Kravtsov, Y. A. & Tatarskii, V. I. *Principles of Statistical Radiophysics* Vol. 3, Ch. 3 (Springer, 1989).
3. Greffet, J.-J. & Nieto-Vesperinas, M. *J. Opt. Soc. Am. A* **15**, 2735–2744 (1998).
4. Hesketh, P. J., Zemel, J. N. & Gebhart, B. *Nature* **324**, 549–551 (1986).
5. Rephaeli, E. & Fan, S. *Opt. Express* **17**, 15145–15159 (2009).
6. Greffet, J.-J. *et al.* *Nature* **416**, 61–64 (2002).
7. Drevillon, J. & Ben-Abdallah, P. *J. Appl. Phys.* **102**, 114305 (2007).
8. Lee, B. J., Fu, C. J. & Zhang, Z. M. *Appl. Phys. Lett.* **87**, 071904 (2005).
9. Schuller, J. A., Taubner, T. & Brongersma, M. L. *Nature Photonics* **3**, 658–661 (2009).
10. Puscasu, I. & Schaich, W. L. *Appl. Phys. Lett.* **92**, 233102 (2008).

CANCER

Why tumours eat tryptophan

Tumours increase their consumption of the amino acid tryptophan to evade immune control. But how does this work? A study shows that the main product of this consumption binds to a receptor involved in the immune system. [SEE ARTICLE P.197](#)

GEORGE C. PRENDERGAST

There is mounting evidence that fast-growing 'progressive' cancers occur because of a failure of the immune system to maintain control over budding tumours. The ability of cancers to escape immune responses is therefore attracting increasing attention, with numerous studies now pointing, perhaps surprisingly, to the consumption of the amino acid tryptophan as a critical factor in progressive cancer. On page 197 of this issue, Opitz *et al.*¹ advance the field with their finding that many cancers upregulate a liver enzyme, tryptophan dioxygenase, to drive tryptophan consumption. What's more, the authors find that the primary product of this process, kynurenine, is an endogenous ligand for the aryl hydrocarbon receptor, which mediates invasive tumour growth. This second finding links the fields of toxicology, immunology and cancer biology in new ways, and may help to explain how elevated tryptophan consumption helps tumours to overcome immune barriers to cancer progression.

Cancer has its roots in altered gene structure and expression, but the tissue microenvironment in which a cancer arises poses a huge barrier to its development and progression. In particular, the complex interplay between cancer cells and nearby immune cells may be one of the most important determinants of whether an early cancer is destroyed by the

immune system, persists in a dormant or slow-growing state (such tumours are often localized and treatable), or progresses to a clinically challenging invasive or metastatic state.

Immune escape represents a budding tumour's victory in its constant thrust and parry with the immune system, which detects accumulating genetic damage in cancer cells. This battle, which is coordinated with oncogenic processes, selects for cancer cells that can resist immune detection or eradication. Immune escape therefore drives the development of tumours towards increasingly aggressive forms². Intriguingly, many pathways of immune escape involve active suppression of the immune system by tumour cells, implying that disruption of these suppressive pathways could restore immune attack. One such pathway that is of emerging importance in cancer involves the consumption of tryptophan and generation of kynurenine by the indoleamine 2,3-dioxygenase enzymes IDO and IDO2, and also by tryptophan dioxygenase (TDO), as Opitz *et al.*¹ report.

The fundamental role of IDO in immune tolerance was recognized several years before its connections to cancer, and it is by far the most broadly expressed and studied of the tryptophan-metabolizing enzymes^{3,4}. IDO's action leads to tryptophan depletion and kynurenine generation, which cooperate to inhibit the activation of immune cells known as T cells through various mechanisms that also affect the activities of other classes of

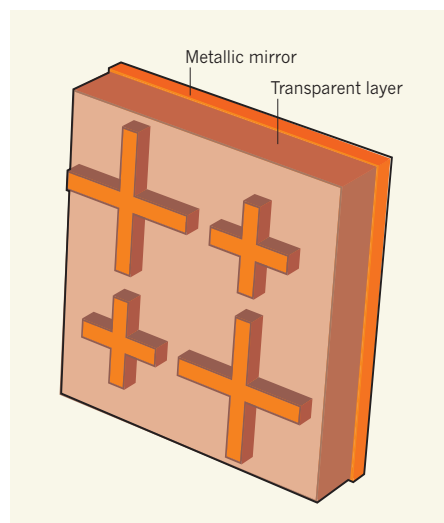


Figure 1 | Infrared thermal emitter. Liu *et al.*¹ have designed a periodic material that can emit infrared light at two wavelengths. The unit cell of the material, represented here, contains two types of cross, each type having a different arm length. The crosses are separated from a metallic mirror by a thin (0.3-micrometre) transparent layer. (Modified from ref. 1.)

immune cells^{5,6}. In addition to cancer cells themselves, dendritic cells and regulatory B cells of the immune system may be key sites of IDO action for initiating and maintaining antigenic tolerance, the process that prevents the immune system from recognizing and attacking tumours. IDO has also been implicated in chronic viral infections and allergies, and in various autoimmune and inflammatory disorders in which immune control is disrupted⁷.

In patients with cancer, the upregulation of IDO is often associated with a poor prognosis⁸. Genetic ablation of the enzyme in mice has confirmed that IDO has crucial roles in immune tolerance and inflammatory carcinogenesis (tumour growth associated with chronic inflammation), with IDO expression in tumour cells, immune cells and stromal cells apparently contributing to its functions in cancer-associated inflammation, immune escape and tumour outgrowth^{9,10}. Pharmacological studies have shown that IDO inhibitors are efficacious anticancer agents, especially when used as adjuvants to improve the efficacy of immunochemotherapy, radiotherapy and cancer vaccines¹¹. Indeed, clinical trials of IDO inhibitors as adjuvant anticancer agents are now under way. What's more, the powerful efficacy of the anticancer drug imatinib in the treatment of solid gastrointestinal stromal tumours may be a result of IDO inhibition by the drug¹².

Opitz *et al.*¹ reveal that, like IDO, TDO is frequently activated in cancer, particularly in cases in which IDO is not activated. This is the situation in a type of brain cancer called glioma, for example, in which the authors performed their studies. Their work is the first to connect TDO to cancer biology. One unique aspect of TDO is that its activation promotes cancer-cell migration, something that IDO has not been reported to do. This suggests some divergence in function between the two enzymes, despite their shared ability to generate kynurenine. TDO is structurally dissimilar to IDO and IDO2, but all three enzymes can consume substrates other than tryptophan. If TDO's substrate preference differs from that of the IDO enzymes, this might differentiate its biological functions from those of IDO or IDO2 to some extent. Whatever the case, Opitz and colleagues' work suggests that TDO inhibitors might be important for cancer studies, both because they may be useful in treating IDO-independent cancers and because TDO activation could be one way for tumours to acquire resistance to IDO inhibitors.

Notably, Opitz *et al.* also report that kynurenine is an endogenous ligand of the aryl hydrocarbon receptor (AHR) that mediates a signalling pathway from TDO (or IDO) to AHR in driving malignant growth (Fig. 1). AHR is a xenobiotic receptor — one that responds to foreign substances in the body — and is most widely known to bind to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, a potent

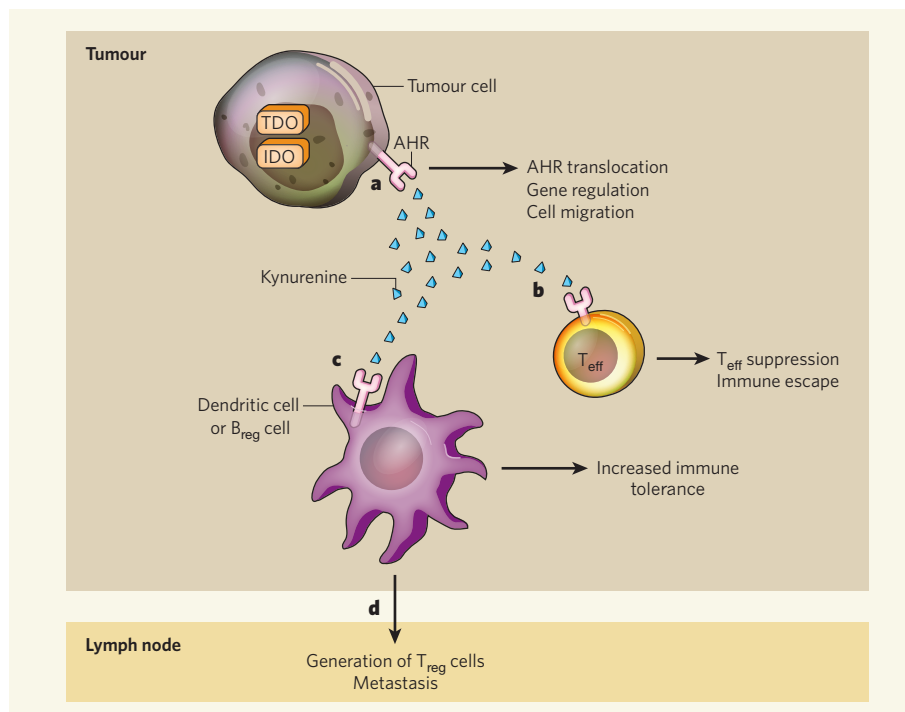


Figure 1 | The kynurenine pathway of cancer invasion and immune escape. Indoleamine 2,3-dioxygenase (IDO) enzymes generate the compound kynurenine in tumour and immune cells. Opitz *et al.*¹ find that another enzyme, tryptophan dioxygenase (TDO), also generates kynurenine in tumour cells, and that kynurenine acts as a signal for the aryl hydrocarbon receptor (AHR). **a**, Kynurenine generated by a tumour cell may activate AHRs on the same cell, causing the receptors to translocate to the cell's nucleus (not shown), where they regulate the expression of target genes that help tumour cells migrate. **b**, Binding of kynurenine to AHRs on certain immune cells (effector T cells; T_{eff}) that recognize tumour cells suppresses the activity of the T_{eff} cells, helping tumours to evade immune responses. **c**, The binding of kynurenine to other tumour-targeting immune cells — dendritic cells and regulatory B (B_{reg}) cells — increases immune tolerance. **d**, Immune tolerance can spread from the tumour's microenvironment when dendritic cells and B_{reg} cells move to lymph nodes and stimulate the generation of tumour-promoting regulatory T cells, thereby facilitating metastasis.

toxic chemical that powerfully suppresses antibody and cellular immune responses, stimulates carcinogenesis and promotes tumour outgrowth. Previous work has suggested that kynurenine can bind to AHR to promote the generation of immune-suppressive T cells that support cancer development¹³. The binding of kynurenine to AHR causes the receptor to move into the nucleus, where it binds target genes and activates their transcription, leading to tumour migration. Genetic studies in mice have defined roles for AHR in immune regulation, inflammation and carcinogenesis¹⁴, the same processes in which IDO has been implicated¹⁵.

Opitz *et al.*¹ found that elevated levels of AHR in tumours corresponded with a poor prognosis in patients, supporting a similar connection between TDO overexpression and patient outcome. Their discovery of kynurenine as an endogenous ligand for AHR in cancer helps to explain why tumours select for IDO- or TDO-mediated tryptophan consumption: the kynurenine that is produced binds to AHR and helps tumours to achieve immune escape. The connection between tryptophan consumption and AHR activation may also

explain why inflammation and immune escape supported by tryptophan consumption are so integrally connected in cancer¹⁵. Intriguingly, although IDO2 is still little understood, the gene encoding it is one of those regulated in dendritic cells by AHR activation¹⁶. Along with other evidence¹⁷, this connection hints at a feedback loop between AHR and tryptophan consumption, suggesting that the IDO/TDO-kynurenine-AHR signalling pathway might dynamically modulate inflammation and immunity.

Given that AHR is likely to be a widespread mediator of kynurenine responses, AHR studies from the field of toxicology could have a large conceptual impact on tumour immunology and cancer biology. Opitz and colleagues' work¹ also begs many questions. Does kynurenine act as a general migratory signal for immune cells or cancer cells? Might AHR effector signals distinguish between the functions of IDO and TDO in inflammation and immune tolerance? And could existing AHR activators or blockers be used to treat diseases that involve IDO, TDO and kynurenine? These issues provide a basis for exciting research opportunities in the future. ■

George C. Prendergast is in the Lankenau Institute for Medical Research, Wynnewood, Pennsylvania 19096, USA.
e-mail: prendergast@limr.org

- Opitz, C. A. *et al.* *Nature* **478**, 197–203 (2011).
- Schreiber, R. D., Old, L. J. & Smyth, M. J. *Science* **331**, 1565–1570 (2011).
- Munn, D. H. & Mellor, A. L. *J. Clin. Invest.* **117**, 1147–1154 (2007).
- Prendergast, G. C. *Oncogene* **27**, 3889–3900 (2008).
- Mellor, A. L. & Munn, D. H. *Nature Rev. Immunol.*

- 8**, 74–80 (2008).
- Fallarino, F., Gizzi, S., Mosci, P., Grohmann, U. & Puccetti, P. *Curr. Drug Metab.* **8**, 209–216 (2007).
- Katz, J. B., Muller, A. J. & Prendergast, G. C. *Immunol. Rev.* **222**, 206–221 (2008).
- Uytendhoeve, C. *et al.* *Nature Med.* **9**, 1269–1274 (2003).
- Muller, A. J., Mandik-Nayak, L. & Prendergast, G. C. *Immunotherapy* **2**, 293–297 (2010).
- Muller, A. J. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 17073–17078 (2008).
- Muller, A. J., DuHadaway, J. B., Donover, P. S., Sutanto-Ward, E. & Prendergast, G. C. *Nature Med.* **11**, 312–319 (2005).

- Balachandran, V. P. *et al.* *Nature Med.* **17**, 1094–1100 (2011).
- Mezrich, J. D. *et al.* *J. Immunol.* **185**, 3190–3198 (2010).
- Stevens, E. A., Mezrich, J. D. & Bradfield, C. A. *Immunology* **127**, 299–311 (2009).
- Prendergast, G. C., Metz, R. & Muller, A. J. *Am. J. Pathol.* **176**, 2082–2087 (2010).
- Vogel, C. F. A., Goth, S. R., Dong, B., Pessah, I. N. & Matsumura, F. *Biochem. Biophys. Res. Commun.* **375**, 331–335 (2008).
- Nguyen, N. T. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 19961–19966 (2010).

EARTH SCIENCE

Sea change for the rise of oxygen

A model proposes that falling sea levels shifted the make-up of volcanic gases on the early Earth, triggering a chain of events that may have allowed photosynthesis in the ocean to oxygenate the atmosphere. SEE LETTER P.229

TIMOTHY W. LYONS &
CHRISTOPHER T. REINHARD

A burning question for those studying the early Earth is whether oxygen production started in the ocean before it first began to accumulate in the atmosphere. A growing body of work convincingly shows that cyanobacteria in the ocean produced oxygen roughly 200 million to 300 million years before the Great Oxidation Event (GOE) — the first appreciable build-up of oxygen in the atmosphere, which occurred about 2.4 billion years ago. But for oxygen to have been produced without accumulating in the atmosphere, efficient sinks or buffers must have existed to consume it as quickly as it was generated. On page 229 of this issue, Gaillard and co-authors¹ offer their take on this conundrum. They propose that falling sea levels across growing, stabilizing continents altered the gas composition of volcanoes, and that these tectonic changes weakened oxygen buffers, enabling the GOE.

The idea of pre-GOE oxygen production received a big boost with the recognition² that organic molecules preserved in roughly 2.7-billion-year-old shaley rocks in Western Australia are biomarkers both of cyanobacterial oxygen production and of oxygen-demanding steroid biosynthesis by eukaryotes (organisms such as plants, animals and fungi). The validity of these biomarkers has since been questioned^{3,4}, but other strong evidence has been found indicating that oxygen was produced⁵, and at times accumulated at very low levels^{6–8}, before the GOE.

Previous models of oxygen buffering

postulated the release of chemically reduced volcanic gases, such as hydrogen, that reacted efficiently with oxygen emerging from biological sources. Only when the levels of those oxygen-consuming gases fell did atmospheric oxygenation begin, producing the clear, irreversible evidence now thought of as the smoking gun of the GOE — the permanent

disappearance of signatures known as mass-independent fractionations (MIFs) from the sulphur isotopic record of sedimentary minerals such as pyrite⁹.

A critical subtheme that has gained traction in oxygen-buffering models is the coincident, but probably not coincidental, occurrence of the GOE with one of Earth's great tectonic transitions — a purported spurt in the generation and stabilization of continents and their emergence from the sea. In 2007, Lee Kump and Mark Barley¹⁰ noted another important transition that paralleled the rise of oxygen levels and of the continents: a shift from previously dominant submarine volcanoes to subaerial volcanoes — those on land. They argued that the two types of volcanoes had fundamentally different redox states because of differences in the temperature and pressure at which their volatile components degassed from molten rock. Hotter, land volcanism would therefore have resulted in bountiful oxidized gases, including sulphur

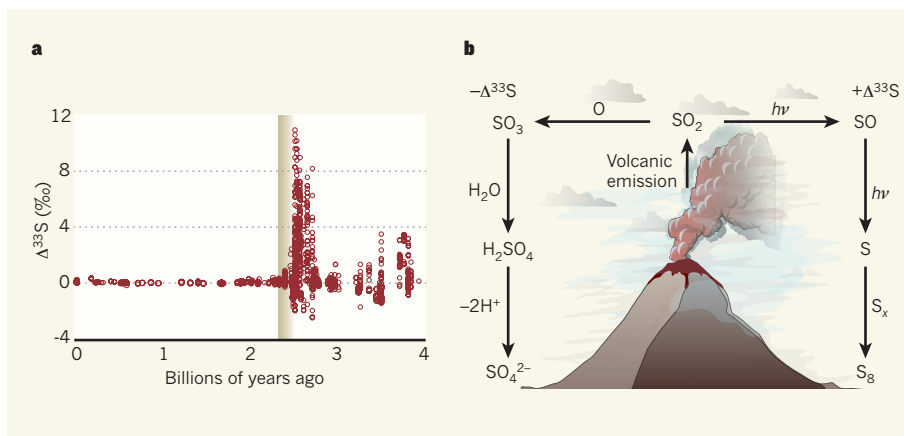


Figure 1 | Sulphur isotopic data through Earth's history. **a**, The isotopic composition of sulphur on Earth can be plotted as Δ³³S — a measure of the deviation of a mineral's sulphur isotopic composition from that expected if the variations scaled strictly with the masses of the isotopes, expressed in parts per thousand (‰). Values above or below 0‰ indicate the magnitude of 'mass-independent fractionation' (MIF), which is linked to atmospheric chemistry and volcanic-gas releases. Before the Great Oxidation Event (shaded box), atmospheric chemical reactions such as those shown in **b** led to large Δ³³S anomalies following the release of sulphur dioxide (SO₂) from volcanoes into the atmosphere. Gaillard *et al.*¹ suggest that these anomalies were caused by an increase in the ratio of SO₂ to hydrogen sulphide (H₂S) in volcanic emissions, and by an increase in the total amount of sulphur-containing gases released, both of which were the result of more widespread subaerial volcanism. They further propose that a concomitant increase in oceanic sulphate (SO₄²⁻) allowed biologically produced oxygen to accumulate in the atmosphere, causing subsequent loss of MIF signals. **b**, This simplified scheme shows MIF source reactions in the atmosphere following volcanic emission of SO₂; *hν* indicates light-dependent reactions. The reaction sequences are abridged and depicted as unidirectional to emphasize the main products — sulphate (SO₄²⁻) and elemental sulphur (S₈) — and pathways from the atmosphere to the ocean.

dioxide (SO₂), which are poor oxygen buffers compared with the chemically reduced gases, such as hydrogen and hydrogen sulphide (H₂S) resulting from the earlier, mainly submarine eruptions. The GOE followed in response.

Gaillard *et al.*¹ now give us a way to think about continents, volcanoes and the GOE that builds on Kump and Barley's model. On the basis of studies of lava from modern Hawaii, the authors propose that submarine lavas need not be more reduced than those above the sea. They also performed thermodynamic calculations suggesting that the properties of volcanic gases — specifically, the relative concentrations of SO₂ and H₂S — are little affected by the redox state of lava.

The authors go on to provide a new spin on the volcano theory based on a model of the gas–melt equilibrium of Hawaiian tholeiitic basalt (which is assumed to be a good analogue for the early Earth). Specifically, they attribute increases in the overall amount of sulphur-containing gases released by volcanoes just before the GOE, and the accompanying major shift from reduced H₂S towards oxidized SO₂, simply to decreases in the average pressure of volcanic degassing. The emergence of continents jutting well above the seas would have produced the ubiquitous subaerial volcanoes that lowered the pressure of eruptions. The argument for continental emergence rests with another model¹¹ asserting that long-term cooling of the mantle allowed for thicker, stronger continental crust, which rose above the oceans about 2.5 billion years ago.

Gaillard and colleagues' theory¹ is built largely on theoretical thermal and chemical models that have ample degrees of uncertainty. But another key part of their proposition relies on data: the structure of sulphur's MIF record. The disappearance of MIF signals for sulphur in sedimentary minerals mirrors the rise of atmospheric oxygen at the GOE (Fig. 1), but equally impressive is the large magnitude of MIF signals that occur in the mineral record just before the disappearance. A recent model¹² ties these large MIF signals to high ratios of SO₂ to H₂S roughly 2.7 billion to 2.5 billion years ago, but does not focus on the mechanisms behind the changing gas composition.

Gaillard *et al.* posit that the mechanistic key to atmospheric oxygenation might have been increased delivery of sulphate to the ocean through the same atmospheric, light-dependent reactions that yielded the high-magnitude MIF signals (Fig. 1). Simply put, increased inputs of sulphate (and indeed increased inputs of sulphur-containing gases overall) could have ramped up the production of H₂S from sulphate reduction occurring at high-temperature volcanic vents on the sea floor. This H₂S might then have reacted with reduced, dissolved iron released from the vents that would otherwise have served as an oxygen buffer. The greater flux of sulphur

could even have allowed H₂S to accumulate in parts of the ocean¹³.

Questions certainly remain about the buffering systems and the timing of events in Gaillard and colleagues' model. For example, a peak in the abundance of huge sedimentary iron formations occurs 2.7 billion to 2.5 billion years ago — the same time as MIF signals reached their apex. This relationship indicates that large amounts of reduced iron emerged from hydrothermal vents at that time, despite the putative increase of sulphate in the ocean. What's more, the timing of events in thermal (tectonic) models¹¹ fits only loosely with that of the relevant chemical models and with the timing of the transitional atmospheric oxygenation that occurred before the GOE proper^{6–8}. The specific relationships between sulphate availability in the ocean and the processes that buffer iron release from deep-sea vents are open to other interpretations¹⁴ that may challenge the authors' emphasis on thermochemical sulphate reduction and attendant H₂S production. Also, there is much uncertainty about when continental growth and emergence from the ocean occurred.

Gaillard and colleagues' model falls among a growing number of oxygen-buffering scenarios, including one based on molybdenum¹³ that is linked to nitrogen bioavailability in the ocean and another that involves nickel¹⁵ and its role in methane production (methane can be an important oxygen sink). It may well be that many diverse processes

controlled the oxygenation of the atmosphere to varying extents, and that the ties they share to the emergence of continents on a cooling Earth are anything but coincidental. ■

Timothy W. Lyons and Christopher T. Reinhard are in the Department of Earth Sciences, University of California, Riverside, Riverside, California 92521, USA.
e-mails: timothyw@ucr.edu; crein003@ucr.edu

1. Gaillard, F., Scaillet, B. & Arndt, N. T. *Nature* **478**, 229–232 (2011).
2. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. *Science* **285**, 1033–1036 (1999).
3. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. *Nature* **455**, 1101–1104 (2008).
4. Rashby, S. E., Sessions, A. L., Summons, R. E. & Newman, D. K. *Proc. Natl Acad. Sci. USA* **104**, 15099–15104 (2007).
5. Waldbauer, J. R., Sherman, L. S., Sumner, D. Y. & Summons, R. E. *Precamb. Res.* **169**, 28–47 (2009).
6. Anbar, A. D. *et al. Science* **317**, 1903–1906 (2007).
7. Kendall, B. *et al. Nature Geosci.* **3**, 647–652 (2010).
8. Waldbauer, J. R., Newman, D. K. & Summons, R. E. *Proc. Natl Acad. Sci. USA* **108**, 13409–13414 (2011).
9. Farquhar, J., Zerkle, A. L. & Bekker, A. *Photosynth. Res.* **107**, 11–36 (2011).
10. Kump, L. R. & Barley, M. E. *Nature* **448**, 1033–1036 (2007).
11. Flament, N., Coltice, N. & Rey, P. F. *Earth Planet. Sci. Lett.* **275**, 326–336 (2008).
12. Halevy, I., Johnston, D. T. & Schrag, D. P. *Science* **329**, 204–207 (2010).
13. Scott, C. T. *et al. Geology* **39**, 119–122 (2011).
14. Kump, L. R. & Seyfried, W. E. Jr *Earth Planet. Sci. Lett.* **235**, 654–662 (2005).
15. Konhauser, K. O. *et al. Nature* **458**, 750–753 (2009).

QUANTUM ENGINEERING

Spins coupled to a persistent current

Quantum computing architectures based on hybrid systems require strong coupling and information exchange between their constituent elements. These two features have been achieved in one such hybrid setting. SEE LETTER P.221

IRINEL CHIORESCU

The development of electronic technologies based on the principles of quantum mechanics, such as quantum computing, requires coupling and integration of quantum objects of various kinds on the same electronic chip. For such integration to succeed, each object needs to be among the best in its class. On page 221 of this issue, Zhu *et al.*¹ make progress in this direction by successfully demonstrating coupling between two quantum systems: a superconducting flux quantum bit (qubit) and an ensemble of identical, highly 'coherent' quantum spins. The authors show that the two systems exchange

quanta of radiation, allowing quantum information encoded in the quanta to be reliably transferred between them. The result is relevant for quantum computing based on hybrid settings, in which the superconducting qubits would process quantum information and the spins would be used to preserve or transfer the information².

Superconducting flux (or persistent current) qubits are loops of superconducting material interrupted by insulating barriers known as Josephson junctions. They are an excellent choice among superconducting qubits. This is because their magnetic flux, which is produced by the current circulating in the loop, couples directly to quantum magnets (atomic

spins) that either exist in crystals or are artificially induced in them. In Zhu and colleagues' case, the quantum magnets are associated with the electronic spin states of atomic defects created in diamond. The two qubit states, the ground and excited states, are defined respectively by the clockwise and anticlockwise direction of a persistent current of hundreds of nanoamperes. Being a quantum, as opposed to a classical, bit, the qubit can also be in a quantum superposition of the two states — that is, it can be in the ground and excited states simultaneously. In their experiment, Zhu *et al.* used a qubit that has four Josephson junctions³ (Fig. 1), which allow adjustment of the energy difference between the two qubit states to match the energetic fingerprint of the diamond spins, as well as optimal operation of the qubit.

Coupling between atoms and electromagnetic fields in resonators was first studied in atomic physics because atomic energy levels in diluted gases have a sufficient lifetime for the transfer of quanta to be observed. Proposed by Tavis and Cummings⁴ in the 1960s, the coupling between atoms and a standing electromagnetic wave has been observed for an ensemble of atoms⁵ and single atoms⁶ in cavity quantum electrodynamics experiments, in which the photons and atoms interact in a cavity. More recently, trapping molecules above a superconducting microwave resonator has been proposed⁷ as a means to implement cavity quantum electrodynamics on a chip. These studies^{5–7} made use of the electric-field component of an electromagnetic field to coherently exchange photons between the field and the system under study.

Although the magnetic-field component of an electromagnetic field offers much less coupling than the electric-field analogue, magnetic two-level systems (such as the 'up' and 'down' spins in an ensemble of atoms) have much longer lifetimes than their electrical counterparts. This aspect is central to cavity quantum electrodynamics experiments in solid-state systems, which are relevant to the implementation of quantum computing on a chip but rapidly lose quantum information to the surrounding environment. Dipolar interactions between spins in such systems are a major cause of information loss because they affect the spin lifetime. But these interactions can be reduced by 'diluting' the spins in the host solid. Strong magnetic coupling between the electromagnetic field in the cavity and a

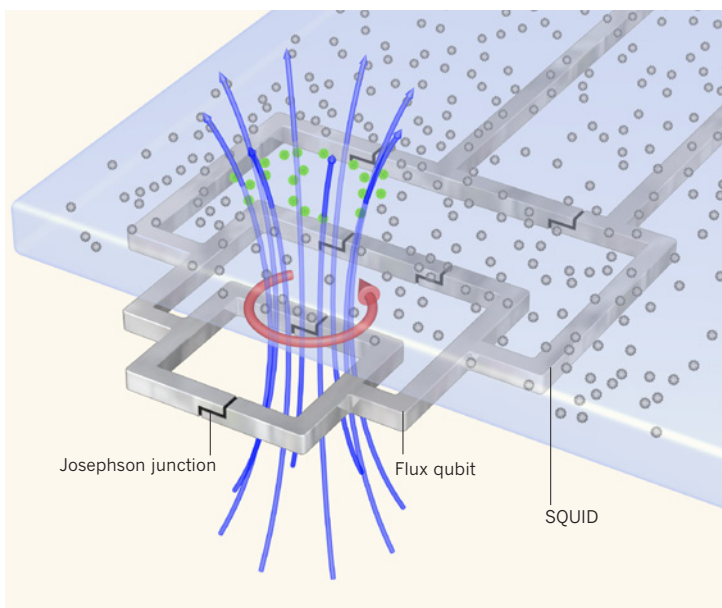


Figure 1 | Quantum hybrid. Zhu and colleagues' hybrid quantum system¹ involves a slab of diamond (blue) placed on top of a superconducting circuit (grey). A persistent current (red), the direction of which can be either clockwise or anticlockwise, characterizes a tunable flux quantum bit featuring four Josephson junctions. A larger structure known as a superconducting quantum interference device (SQUID), which features two additional Josephson junctions, is used to detect the average value of the persistent current. The magnetic flux (blue arrows) generated by the persistent current couples to some of the spins (green dots) in the diamond. The spins (grey and green dots) are in the form of defects located in a plane inside the diamond slab that is separated from the flux qubit by about 1.2 micrometres. This coupling allows quantum information to be transferred from the flux qubit to the spin ensemble.

spin ensemble has been demonstrated theoretically⁸ and experimentally (see ref. 2 for a discussion), and requires a coupling strength larger than both the cavity's photon decay rate and the rate at which the spin loses its quantum-state information.

In typical cavity quantum electrodynamics experiments, the cavity field is a harmonic oscillator and is coupled to an ensemble of one or more two-level systems. But in their experiment, Zhu *et al.*¹ reversed the roles of these elements: they demonstrated strong coupling between the flux qubit (a two-level system) and the ensemble of spins, which is treated as a single harmonic oscillator. The authors prepared a diamond slab with a number of nitrogen-vacancy (NV) defects located in a plane inside the diamond that is separated from the flux qubit by about 1.2 micrometres (Fig. 1). Each NV defect has three states, with the state of lowest energy being separated from the two higher-energy states owing to an internal magnetic field in the crystal (see Fig. 1c of the paper¹). It is this energy-level separation that defines the ground and excited states of the ensemble of NV defects and that allows these states to be entangled with the ground and excited states of the flux qubit.

Zhu *et al.* used a flux qubit with an energy gap between its ground and excited states that can be tuned to have the same energy as the energy gap between the ground and excited

states of the NV ensemble. In this way, the NV ensemble and qubit can be coupled and placed in an entangled state in which the qubit is in its excited state and the NV ensemble is in its ground state. This state can be reversed into a state in which the qubit is in the ground state and the ensemble is in its first excited state. This allows quantum information encoded in the qubit state to be stored in the collection of long-lived NV spins. Zhu and colleagues estimated that the number of NV defects that take part in this process is as large as 3×10^7 , giving a collective qubit–NV-ensemble coupling of 70 megahertz, which is sufficiently large to split the energy of the joint state and its reverse (see Fig. 3b of the paper¹). They showed that a coherent oscillation between these two states lasted for about 20 nanoseconds, and suggested ways to increase this duration.

Compared with set-ups in which a resonator is used instead of a flux qubit², the flux qubit gives a more localized coupling. In addition, the

external magnetic fields used to control the two states of the spin system alter a resonator's frequency and so affect the coupling process, which does not happen for a qubit of such small dimensions. What's more, this type of device has a larger qubit–ensemble coupling⁹ than devices involving resonators, and operates in a frequency range of a few gigahertz, which is practical for low-temperature experiments. Finally, a four-junction flux qubit such as that of Zhu and colleagues can be used for studies involving materials other than diamond, because it can be tuned to be in resonance with spin states over a broad energy range. ■

Irinel Chiorescu is in the Department of Physics and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, USA.
e-mail: ichiorescu@fsu.edu

1. Zhu, X. *et al.* *Nature* **478**, 221–224 (2011).
2. Blencowe, M. *Nature* **468**, 44–45 (2010).
3. Mooij, J. E. *et al.* *Science* **285**, 1036–1039 (1999).
4. Tavis, M. & Cummings, F. W. *Phys. Rev.* **170**, 379–384 (1968).
5. Kaluzny, Y., Goy, P., Gross, M., Raimond, J. M. & Haroche, S. *Phys. Rev. Lett.* **51**, 1175–1178 (1983).
6. Thompson, R. J., Rempe, G. & Kimble, H. J. *Phys. Rev. Lett.* **68**, 1132–1135 (1992).
7. Rabl, P. *et al.* *Phys. Rev. Lett.* **97**, 033003 (2006).
8. Imamoglu, A. *Phys. Rev. Lett.* **102**, 083602 (2009).
9. Marcos, D. *et al.* *Phys. Rev. Lett.* **105**, 210501 (2010).

An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor

Christiane A. Opitz^{1,2*}, Ulrike M. Litzenburger^{1,2*}, Felix Sahm³, Martina Ott^{1,2}, Isabel Tritschler⁴, Saskia Trump⁵, Theresa Schumacher^{1,2}, Leonie Jestaedt⁶, Dieter Schrenk⁷, Michael Weller⁴, Manfred Jugold⁸, Gilles J. Guillemin⁹, Christine L. Miller¹⁰, Christian Lutz¹¹, Bernhard Radlwimmer¹², Irina Lehmann⁵, Andreas von Deimling³, Wolfgang Wick^{1,13} & Michael Platten^{1,2}

Activation of the aryl hydrocarbon receptor (AHR) by environmental xenobiotic toxic chemicals, for instance 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (dioxin), has been implicated in a variety of cellular processes such as embryogenesis, transformation, tumorigenesis and inflammation. But the identity of an endogenous ligand activating the AHR under physiological conditions in the absence of environmental toxic chemicals is still unknown. Here we identify the tryptophan (Trp) catabolite kynurenine (Kyn) as an endogenous ligand of the human AHR that is constitutively generated by human tumour cells via tryptophan-2,3-dioxygenase (TDO), a liver- and neuron-derived Trp-degrading enzyme not yet implicated in cancer biology. TDO-derived Kyn suppresses antitumour immune responses and promotes tumour-cell survival and motility through the AHR in an autocrine/paracrine fashion. The TDO-AHR pathway is active in human brain tumours and is associated with malignant progression and poor survival. Because Kyn is produced during cancer progression and inflammation in the local microenvironment in amounts sufficient for activating the human AHR, these results provide evidence for a previously unidentified pathophysiological function of the AHR with profound implications for cancer and immune biology.

Degradation of Trp by indoleamine-2,3-dioxygenases 1 and 2 (IDO1/2) in tumours and tumour-draining lymph nodes inhibits antitumour immune responses^{1–5} and is associated with a poor prognosis in various malignancies⁶. Inhibition of IDO1/2 suppresses tumour formation in animal models^{1,3} and is currently tested in phase I/II clinical trials in patients with cancer⁷. The relevance of Trp catabolism to human tumour formation and progression is, however, as yet unknown.

TDO degrades Trp to Kyn in human brain tumours

A screen of human cancer cell lines revealed constitutive degradation of Trp and release of high micromolar amounts of Kyn in brain tumour cells, namely glioma cell lines and glioma-initiating cells (GICs), but not human astrocytes (Fig. 1a). IDO1 and IDO2 did not account for the constitutive Trp catabolism in brain tumours (Supplementary Fig. 1a–e). Instead, tryptophan-2,3-dioxygenase (TDO), which is predominantly expressed in the liver (Supplementary Fig. 3a, b) and is believed to regulate systemic Trp concentrations⁶, was strongly expressed in human glioma cells (Supplementary Fig. 1b) and correlated with Kyn release (Fig. 1b). Pharmacological inhibition or knockdown of *TDO* blocked Kyn release by glioma cells, whereas knockdown of *IDO1* and *IDO2* had no effect (Fig. 1c, d and Supplementary Fig. 2a), thus confirming that TDO is the central Trp-degrading enzyme in human glioma cells. In human brain tumour specimens, TDO protein levels increased with malignancy and correlated with the proliferation index (Fig. 1e–g and Supplementary Fig. 2b–d). As described previously⁸, healthy human brain showed only weak TDO staining in neurons (Fig. 1e). TDO

expression was not confined to gliomas but was also detected in other cancer types (Supplementary Fig. 3b, c). Lower Trp concentrations were measured in the serum of patients with glioma (Fig. 1h). These may not have translated into increased Kyn levels (Fig. 1h) because Kyn is taken up by other cells and metabolized to quinolinic acid. Indeed, accumulation of quinolinic acid was detected in TDO-expressing glioblastoma tissue (Fig. 1i and Supplementary Fig. 3d).

Effects of TDO-mediated Kyn release on immune cells

Kyn suppresses allogeneic T-cell proliferation⁹. Allogeneic T-cell proliferation was inversely correlated with Kyn formation by glioma-derived TDO (Fig. 2a and Supplementary Fig. 4a, b). Knockdown of *TDO* in glioma cells (Supplementary Fig. 4c, d) restored allogeneic T-cell proliferation, and the addition of Kyn to the *TDO* knockdown cells prevented the restoration of T-cell proliferation (Fig. 2b). The proliferation of CD4⁺ and CD8⁺ T cells stimulated by the T-cell receptor was inhibited by Kyn in a concentration-dependent manner (Supplementary Fig. 4e). In addition, knockdown of *TDO* resulted in enhanced lysis of glioma cells by alloreactive peripheral blood mononuclear cells (Supplementary Fig. 4f). Finally, decreased infiltration with leukocyte common antigen (LCA)-positive and CD8⁺ immune cells was observed in sections of human glioma with high TDO expression in comparison with those with low TDO expression (Fig. 2c), indicating that Kyn formation by TDO may suppress antitumour immune responses. *In vivo* experiments in immunocompetent mice demonstrated that tumours expressing TDO grew

¹Department of Neurooncology, Neurology Clinic and National Center for Tumor Diseases University Hospital of Heidelberg, 69120 Heidelberg, Germany. ²Experimental Neuroimmunology Unit, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ³Department of Neuropathology, Institute of Pathology, University Hospital of Heidelberg and Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁴Department of Neurology, University Hospital Zurich, 8091 Zurich, Switzerland. ⁵Department for Environmental Immunology, Helmholtz Center for Environmental Research, 04318 Leipzig, Germany. ⁶Department of Neuroradiology, University Hospital of Heidelberg, 69120 Heidelberg, Germany. ⁷Food Chemistry and Toxicology, University of Kaiserslautern, 67663 Kaiserslautern, Germany. ⁸Small Animal Imaging Center, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁹Department of Pharmacology, School of Medical Sciences, University of New South Wales, Sydney, 2052 Australia. ¹⁰Department of Pediatrics, Johns Hopkins University, Baltimore, MD 21287, USA. ¹¹Heidelberg Pharma AG, 68526 Ladenburg, Germany. ¹²Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ¹³Clinical Cooperation Unit Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

*These authors contributed equally to this work.

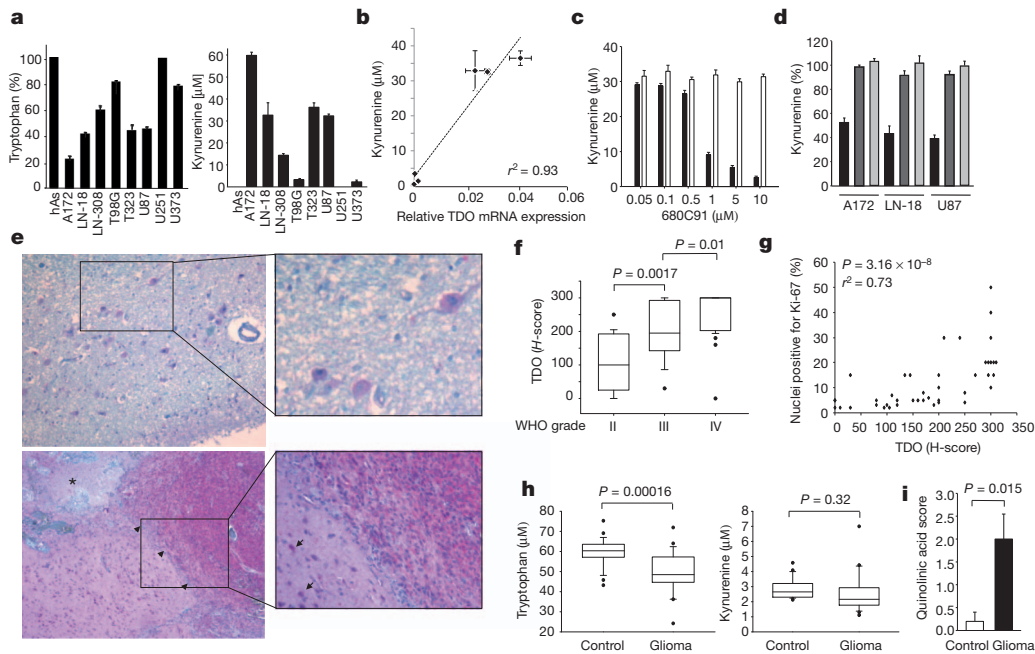


Figure 1 | TDO degrades Trp to Kyn in human brain tumours. **a**, Trp (left) and Kyn (right) content in the supernatants of human astrocytes (hAs), glioma cell lines and GICs (T323) cultured for 72 h and measured by high-performance liquid chromatography (HPLC) ($n = 4$). **b**, Correlation between *TDO* mRNA and Kyn release by human glioma cells measured by quantitative RT-PCR and HPLC ($n = 4$). **c**, Kyn concentrations in the supernatants of U87 glioma cells cultured for 48 h in the presence of the TDO inhibitor 680C91 (black bars) or its solvent (white bars; $n = 4$, $P = 0.005$, 0.002 and 0.0009 for 1, 5 and 10 μM 680C91, respectively). **d**, Kyn release by glioma cells after knockdown of *TDO* (black bars; $P = 0.000007$, 0.00007 and 0.00006, respectively), *IDO1* (dark grey bars) or *IDO2* (light grey bars) by siRNA ($n = 3$). **e**, Upper panel: weak neuronal *TDO* expression in healthy brain tissue. Lower panel: *TDO* expression in glioblastoma (WHO grade IV). *TDO* staining is in red. Asterisk indicates necrosis; arrowheads indicate the border

faster and had a higher proliferation index than *TDO*-deficient control tumours (Fig. 2d and Supplementary Fig. 4g–i). *TDO* activity in tumours suppressed antitumour immune responses *in vivo*, as demonstrated by a decreased release of interferon- γ by tumour-specific T cells and tumour cell lysis by spleen cells of mice bearing *TDO*-expressing tumours in comparison with mice bearing *TDO*-deficient tumours (Fig. 2e, f).

Effects of *TDO*-mediated Kyn release on glioma cells

We next assessed the autocrine effects of Kyn on glioma cells. Although no differences in cell cycle progression were detected between controls and glioma cells with *TDO* knockdown (Supplementary Fig. 5a), knockdown of *TDO* decreased motility and clonogenic survival (Fig. 2g, h and Supplementary Fig. 5b–d). This was mediated by Kyn because exogenous addition of Kyn restored motility and clonogenic survival in the absence of Trp (Fig. 2i, j and Supplementary Fig. 5e, f), suggesting that Kyn increases the motility of malignant glioma cells. In GICs, sphere formation was enhanced in response to Kyn (Supplementary Fig. 5g). Finally, tumour formation was impaired when *TDO* knockdown tumours were orthotopically implanted in the brains of nude mice, which are devoid of functional T cells (Fig. 2k and Supplementary Fig. 5h, i). To analyse whether *TDO*-mediated inhibition of antitumour natural killer (NK)-cell responses, which are functional in nude mice, might account for the impaired formation of *TDO* knockdown tumours, we compared subcutaneous tumour growth in the presence or absence of NK cells. NK-cell depletion (Supplementary Fig. 5j) enhanced the growth of both control and *TDO* knockout tumours but did not restore the growth of *TDO* knockout tumours to that in controls (Fig. 2l and Supplementary Fig. 5k), suggesting that Kyn generated by constitutive *TDO* activity

to infiltrated brain tissue. Insets: single tumour cells (arrows) infiltrating the adjacent brain tissue. Magnifications: $\times 40$ (main panels), $\times 400$ (upper inset) and $\times 100$ (lower inset). **f**, Plot of *TDO* expression (*H*-score; see Supplementary Methods) in brain tumours of increasing malignancy (WHO grades II–IV; grade II, $n = 18$; grade III, $n = 15$; grade IV, $n = 35$). **g**, Correlation of the Ki-67 proliferative index with the *TDO* *H*-score in gliomas of different WHO grades ($n = 42$). **h**, Trp (left) and Kyn (right) concentrations in the sera of 24 patients with glioblastoma and 24 age-matched and sex-matched healthy controls, measured by HPLC. **i**, Quantification of staining with quinolinic acid in healthy human brain tissue (white bar; $n = 5$) and glioblastoma tissue (black bar; $n = 5$). The data distribution in **f** and **g** is presented as box plots, showing the 25th and 75th centiles together with the median; whiskers represent the 10th and 90th centiles, respectively. Error bars indicate s.e.m.

enhances the malignant phenotype of human gliomas in an autocrine manner in the absence of functional antitumour T-cell and NK-cell responses.

Kyn activates the aryl hydrocarbon receptor

For a better understanding of the molecular mechanisms underlying the autocrine effects of Kyn on glioma cells, we performed microarray analyses of Kyn-treated glioma cells revealing broad induction of aryl hydrocarbon receptor (AHR) response genes by Kyn (Fig. 3a and Supplementary Figs 6a, b and 7). Pathway analyses showed that the 25 genes that were most strongly induced by Kyn treatment in U87 cells at 8 h and at 24 h were all directly or indirectly regulated by the AHR (Fig. 3a and Supplementary Fig. 6b). The AHR is a transcription factor of the basic helix–loop–helix (bHLH) Per–Arnt–Sim (PAS) family, which is activated by xenobiotics such as benzo[*a*]pyrene and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD)¹⁰. Malignant glioma cell lines and GICs express the AHR constitutively (Supplementary Fig. 6c)¹¹, and upregulation of AHR target genes by Kyn was confirmed in two different glioma cell lines (Supplementary Fig. 6d, e). Kyn led to translocation of the AHR into the nucleus after 1 h, thus showing an immediate effect of Kyn on the AHR (Fig. 3b, c and Supplementary Fig. 8a). In accordance with this, western blot analyses of Kyn-activated tumour cells showed decreased cytoplasmic localization paralleled by increased nuclear accumulation of the AHR comparable to that induced by TCDD (Fig. 3d). In the nucleus the AHR forms a heterodimer with the AHR nuclear translocator (ARNT) that interacts with the core-binding motif of the dioxin-responsive elements (DRE) located in regulatory regions of AHR target genes¹². Kyn induced DRE–luciferase activity in glioma cells, with a concentration giving half-maximal response (EC_{50}) of 36.6 μM (Fig. 3e,

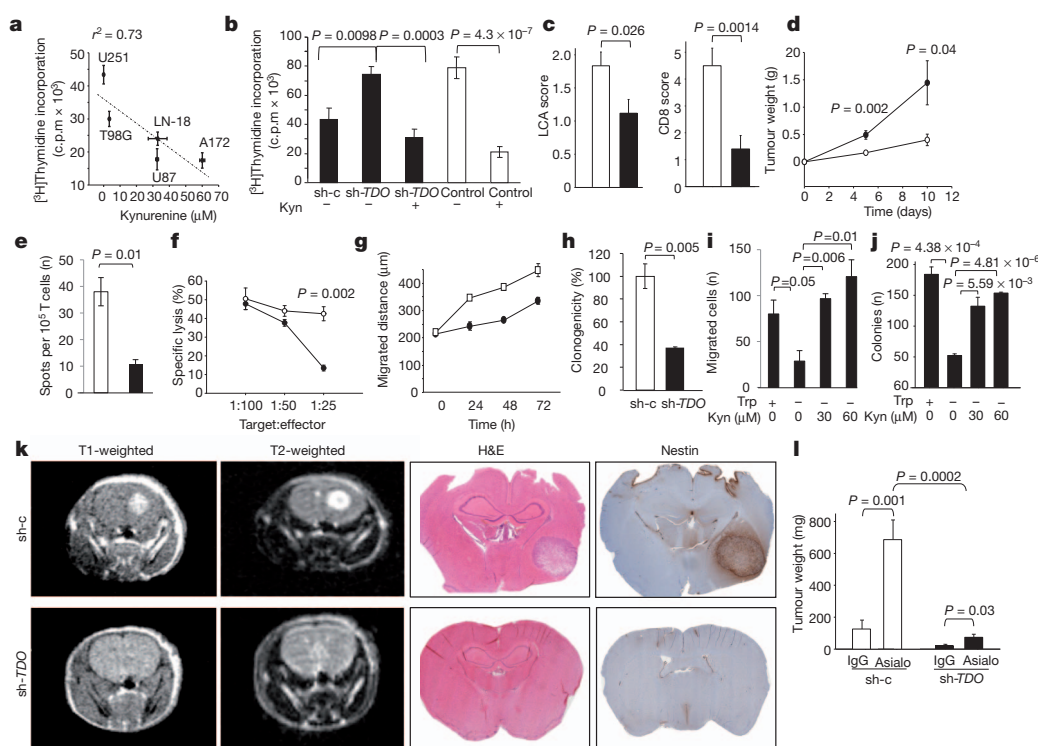


Figure 2 | Paracrine and autocrine effects of TDO-mediated Kyn release by glioma cells. **a**, Correlation of the proliferation of peripheral blood mononuclear cells (PBMCs) cultured with allogeneic glioma cell lines with the Kyn release of the glioma cells ($n = 3$). **b**, Proliferation of PBMCs cultured with allogeneic *TDO*-expressing control U87 glioma cells (sh-c) in comparison with U87 glioma cells with a stable short hairpin RNA-mediated knockdown of *TDO* (sh-*TDO*), with or without 100 μ M Kyn (black bars), in comparison with PBMCs alone with or without 100 μ M Kyn (white bars; $n = 3$). **c**, Quantification of LCA⁺ cells (left) and CD8⁺ cells (right) stained in human glioma sections with low *TDO* expression (*H*-score < 150, white bar; $n = 12$ for LCA, $n = 10$ for CD8) and in human glioma sections with high *TDO* expression (*H*-score ≥ 150 , black bar; $n = 17$ for LCA, $n = 10$ for CD8). **d**, Growth of *Tdo*-deficient GL261 murine glioma cells stably transfected with *Tdo* (filled circles) or empty vector (open circles) injected subcutaneously into the flank of C57BL/6N mice was monitored using metric calipers ($n = 6$). Tumour weight (g) was calculated as $0.5 \times \text{length (cm)} \times \text{width}^2 \text{ (cm}^2\text{)}$. **e**, Interferon- γ release by T cells of mice bearing subcutaneous *Tdo*-expressing tumours (black bar) in comparison with T cells of mice bearing *Tdo*-deficient

tumours (white bar) after re-stimulation with glioma lysates measured by ELISpot ($n = 3$). **f**, Lysis of GL261 murine glioma cells by spleen cells of mice with *Tdo*-expressing GL261 tumours in comparison with those with subcutaneous *Tdo*-deficient GL261 tumours, measured by chromium release ($n = 4$). **g**, Quantification of the migrated distances of sh-c (open squares) and sh-*TDO* (filled circles) cells into a collagen matrix ($n = 3$, $P = 0.004$, 0.0005 and 0.01 for 24, 48 and 72 h, respectively). **h**, Clonogenic survival of sh-c (white bar) and sh-*TDO* (black bar) U87 cells ($n = 3$). **i**, Matrigel Boyden chamber assay of U87 glioma cells in the absence or presence of 70 μ M Trp without or with 30 or 60 μ M Kyn ($n = 3$). **j**, Clonogenic survival of LN-18 glioma cells in the absence or presence of 70 μ M Trp without or with 30 or 60 μ M Kyn ($n = 3$). **k**, Representative cranial MRIs, haematoxylin/eosin staining (H&E) and nestin staining of *CD1 nul/nul* mice implanted with sh-c (upper panels) or sh-*TDO* (lower panels) U87 glioma cells. The images are representative of two independent experiments ($n = 6$). **l**, Tumour weight of sh-c (white bars) and sh-*TDO* (black bars) U87 glioma cells injected subcutaneously in the flank of *CD1 nul/nul* mice that were treated either with anti-asialo GM1 antibody (asialo) for NK-cell depletion or control IgG (IgG) ($n = 8$). Error bars indicate s.e.m.

Supplementary Fig. 8b). AHR activation was unique to Kyn in a panel of Trp catabolites (Supplementary Table 1). An ethoxyresorufin-O-deethylase (EROD) assay confirmed the induction of the functional AHR target gene *CYP1A1*, encoding cytochrome P₄₅₀, family 1, subfamily A, polypeptide 1, with an EC₅₀ of 12.3 μ M for Kyn (Supplementary Fig. 8c). Radioligand binding assays with mouse liver cytosol from *Ahr*-proficient and *Ahr*-deficient mice showed that Kyn binds to the AHR with an apparent K_d of roughly 4 μ M (Fig. 3f).

Activation of the AHR and upregulation of AHR-regulated gene expression in response to Kyn were inhibited by the AHR antagonist dimethoxyflavone or knockdown of *AHR* (Fig. 3g and Supplementary Fig. 8g–k), indicating that Kyn is a specific agonist of the AHR. The involvement of the same or similar AHR residues in the binding to Kyn, TCDD and 3-methylcholanthrene was confirmed by the fact that dimethoxyflavone inhibited the activation of the AHR by all three ligands (Supplementary Fig. 8g–i).

The effects of TDO-derived Kyn are mediated by the AHR

The endogenous production of Kyn in glioma cells was sufficient to activate the AHR, because knockdown of *TDO* decreased the expression of AHR-regulated genes (Fig. 3h and Supplementary Fig. 8l–o).

Because mean Kyn concentrations of $37.01 \pm 13.4 \mu$ M were measured in U87 xenografts ($n = 6$), sufficient Kyn concentrations to activate the AHR were also reached *in vivo*.

In accordance with activation of the AHR by TDO-derived Kyn, expression of the AHR target gene *TIPARP* in LCA⁺ immune cells was observed only in human glioma sections expressing *TDO* (Fig. 4a). To determine whether *TDO* influences antitumour immune responses through the AHR we analysed the infiltration of immune cells in human glioma sections in relation to their AHR expression. Indeed, infiltration by LCA⁺ and CD8⁺ immune cells was decreased in sections of human gliomas with high AHR expression compared with those with low AHR expression (Fig. 4b). To analyse the contribution of host AHR expression to tumour growth, we compared the growth of murine tumours with and without *Tdo* expression in *Ahr*-deficient and *Ahr*-proficient mice. The growth of *Tdo*-expressing tumours was attenuated in *Ahr*-deficient mice when compared with *Ahr*-proficient mice (Fig. 4c) indicating that AHR-mediated host effects enhance tumour growth. Staining of LCA⁺ immune cells in the tumours revealed that expression of *TDO* decreased the infiltration with LCA⁺ immune cells in *Ahr*-proficient mice but not in *Ahr*-deficient mice (Fig. 4d and Supplementary Fig. 9a), suggesting that

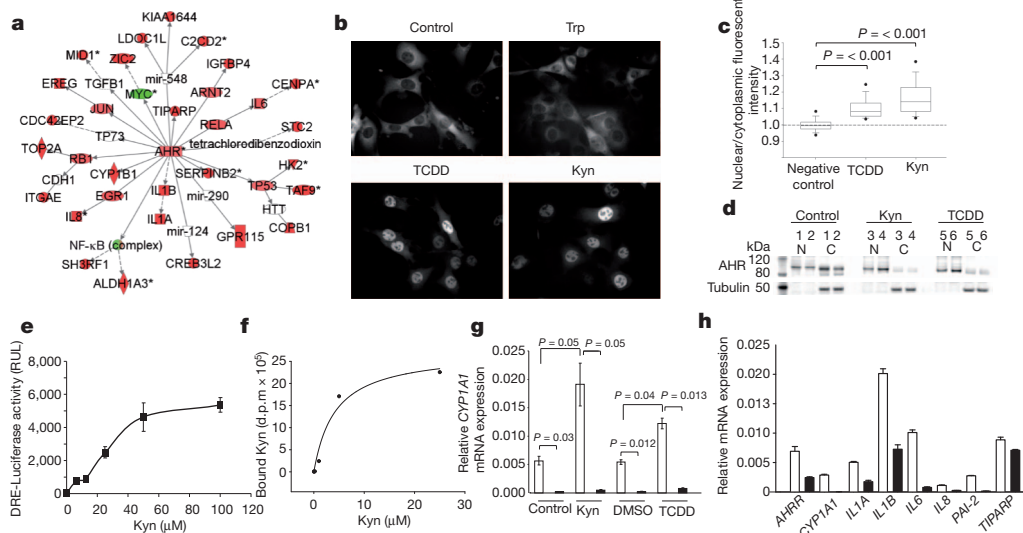


Figure 3 | Kyn activates the AHR. **a**, Connection of the 25 genes that were most strongly induced by Kyn treatment in U87 cells after 8 h to AHR signalling (red, upregulation; green, downregulation). **b**, Translocation of green fluorescent protein (GFP)-tagged AHR into the nucleus of mouse hepatoma cells, which do not degrade Trp, after 3 h treatment with 50 μ M Kyn, 50 μ M Trp or 1 nM TCDD (negative control: medium). **c**, Ratios of nuclear to cytoplasmic fluorescent intensity in cells with GFP-tagged AHR after 3 h of indicated treatment (negative control: medium; positive control: 1 nM TCDD, 50 μ M Kyn). The data distribution is represented by box plots, showing the 25th and 75th centiles together with the median; whiskers represent the 10th and 90th centiles, respectively ($P < 0.001$, one-way analysis of variance on ranks, followed by Dunn's method). **d**, AHR western blots of two different nuclear (N) and cytoplasmic (C) fractions each of control (lanes 1 and 2), Kyn-treated (lanes 3

and 4) and TCDD-treated (lanes 5 and 6) human LN-229 glioma cells. **e**, Dioxin-responsive element (DRE) chemical activated luciferase gene expression in U87 glioma cells treated with the indicated Kyn concentrations ($n = 2$). **f**, Radioligand binding assay with indicated concentrations of L-[3 H]Kyn using mouse liver cytosol from *Ahr*-proficient and *Ahr*-deficient mice. Specific binding was calculated by subtracting the radioactivity measured in *Ahr*-deficient cytosol from that of *Ahr*-proficient cytosol ($n = 4$). **g**, *CYP1A1* mRNA expression in sh-*AHR* LN-308 glioma cells (black bars) in comparison with controls (sh-c; white bars) treated with 100 μ M Kyn, 1 nM TCDD or controls ($n = 4$). **h**, mRNA expression of AHR target genes in sh-*TDO* (black bars) in comparison with sh-c U87 glioma cells (white bars; $n = 4$) (*AHR*, $P = 0.02$; *CYP1A1*, $P = 0.0007$; *IL1A*, $P = 0.001$; *IL1B*, $P = 0.0000006$; *IL6*, $P = 0.0047$; *IL8*, $P = 0.01$; *PAI-2*, $P = 0.0005$; *TIPARP*, $P = 0.06$). Error bars indicate s.e.m.

TDO-mediated suppression of antitumour immune responses through the AHR contributes to the host effects enhancing the growth of *Tdo*-expressing tumours. In addition, while in *Ahr*-proficient mice the expression of *Tdo* strongly enhanced tumour growth in comparison with tumours not expressing *Tdo*, the same effect was

observed in *Ahr*-deficient mice, although to a much smaller extent (Fig. 4c). Because murine glioma cells express functional AHR (Supplementary Fig. 9b–e), these results suggest that the increase in tumour growth mediated by TDO in *Ahr*-deficient mice is due to autocrine effects of TDO on the tumour cells themselves.

This notion is supported by the fact that Kyn failed to induce motility of human glioma cells after *AHR* knockdown (Fig. 4e). In addition, the increase in clonogenic survival in response to Kyn was abolished in glioma cells with a knockdown of *AHR* (Fig. 4f). Finally,

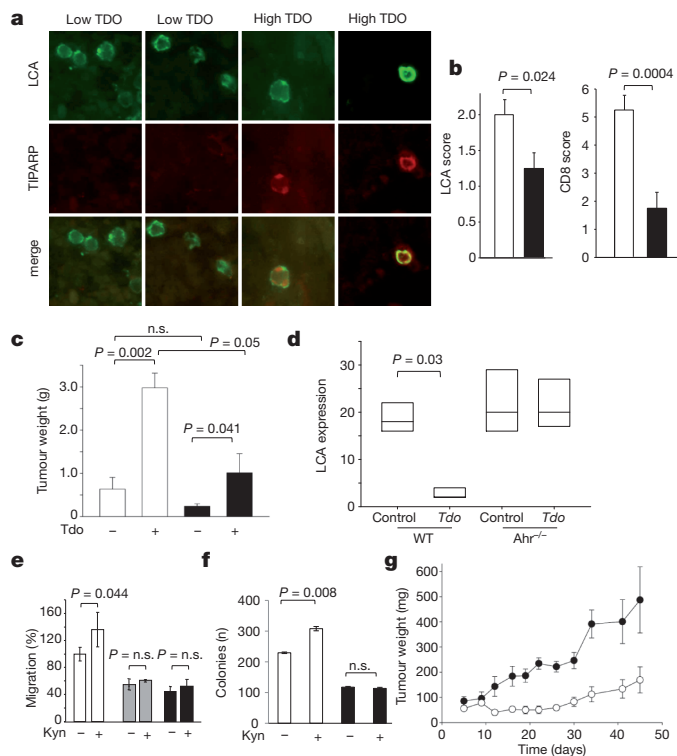


Figure 4 | The autocrine and paracrine effects of TDO-derived Kyn are mediated through the AHR. **a**, Immunofluorescence staining of LCA and TIPARP in human glioma sections with low or high TDO expression. Magnification $\times 400$. **b**, Quantification of LCA $^+$ cells (left) and CD8 $^+$ cells (right) stained in human glioma sections with low AHR expression (HistoScore < 150 , white bar; $n = 10$ for LCA, $n = 8$ for CD8) and in human glioma sections with high AHR expression (H-score ≥ 150 , black bar; $n = 12$ for LCA, $n = 12$ for CD8). **c**, Tumour weight measured 15 days after subcutaneous injection of murine GL261 glioma cells with and without *Tdo* expression in the flanks of *Ahr*-proficient mice (white bars) or *Ahr*-deficient mice (black bars; $n = 6$). **d**, Quantification of LCA $^+$ immune cells stained in the subcutaneous *Tdo*-proficient and *Tdo*-deficient GL261 tumours in *Ahr*-proficient and *Ahr*-deficient mice presented as box plots, showing the 25th and 75th centiles and the median ($n = 4$). WT, wild type. **e**, Migration of sh-c LN-308 glioma cells (white bars) and LN-308 glioma cells with knockdown of the *AHR* by two different shRNAs (grey bars, sh-*AHR1*; black bars, sh-*AHR2*) in the presence or absence of 100 μ M Kyn ($n = 4$). **f**, Clonogenicity of sh-c (white bars) and sh-*AHR* (black bars) LN-308 glioma cells with or without 100 μ M Kyn ($n = 3$). **g**, Growth of *AHR*-proficient (filled circles) and *AHR*-deficient (open circles) human LN-308 glioma cells injected subcutaneously into the flank of *CD1nu/nu* mice with metric callipers ($n = 7$). Tumour weight was calculated as in Fig. 2d. P values: day 5, 0.147; day 9, 0.546; day 12, 0.027; day 16, 0.008; day 19, 7.18×10^{-4} ; day 22, 1.77×10^{-5} ; day 26, 1.57×10^{-5} ; day 30, 8.49×10^{-4} ; day 34, 8.26×10^{-4} ; day 41, 0.022; day 45, 0.0477. Error bars indicate s.e.m.

in vivo experiments demonstrated that induced knockdown of AHR in human glioma cells inhibited tumour growth in immunocompromised mice (Fig. 4g and Supplementary Fig. 9f), underscoring the importance of AHR signalling for the autocrine effects of Trp degradation.

TDO-derived Kyn activates the AHR in human cancer

Next we aimed at investigating whether TDO-derived Kyn activates the AHR in human brain tumour tissue. Indeed, TDO expression correlated with the expression of the AHR and AHR target genes in human glioma tissue (Fig. 5a–c and Supplementary Fig. 9g), indicating that constitutive TDO expression in glioma cells produced sufficient Kyn levels to activate the AHR. To address whether the TDO–Kyn–AHR signalling pathway is also activated in cancers other than glioma, we analysed microarray data of diverse human tumour entities. Interestingly, TDO expression correlated with the expression of the AHR target gene *CYP1B1* not only in glioma (Fig. 5c) but also in B-cell lymphoma, Ewing sarcoma, bladder carcinoma, cervix carcinoma, colorectal carcinoma, lung carcinoma and ovarian carcinoma (Fig. 5d, Supplementary Fig. 10a and Supplementary Table 2). This finding indicates that the TDO–Kyn–AHR pathway is not confined to brain tumours but seems to be a common trait of cancers. Analysis of the Rembrandt database revealed that the overall survival of patients with glioma (WHO grades II–IV) with high expression of TDO, the AHR or the AHR target gene *CYP1B1* was reduced in comparison with patients with intermediate or low expression of these genes (Fig. 5e and Supplementary Fig. 10b)¹³. Finally, in patients with glioblastoma (WHO grade IV)¹⁴, the expression of the AHR targets *CYP1B1*, *IL1B*, *IL6* and *IL8*, which are regulated by TDO-derived Kyn in glioma cells (Fig. 3h and Supplementary Fig. 5d, e), were found to predict survival independently of WHO grade (Fig. 5f and Supplementary Fig. 10c), thus further underscoring the importance of AHR activation for the malignant phenotype of gliomas. In summary, these data suggest that endogenous tumour-derived Kyn

activates the AHR in an autocrine/paracrine fashion to promote tumour progression (Fig. 5g).

Discussion

Cancer-associated immunosuppression by Trp degradation has until now been attributed solely to the enzymatic activity of IDO in cancer cells and tumour-draining lymph nodes. IDO inhibition is therefore currently being evaluated as a therapeutic strategy in the treatment of cancer in clinical trials⁷ despite some off-target effects on human cancer cells¹⁵. We show that TDO is strongly expressed in cancer and is equally capable of producing immunosuppressive Kyn. In IDO-negative glioma cells, TDO seems to be the sole determinant of constitutive Trp degradation, indicating that TDO represents a novel therapeutic target in glioma therapy. In fact, an orally available TDO inhibitor has recently been developed¹⁶. Inhibition of TDO may not only restore antitumour immune responses but also act on the tumour cell intrinsic malignant phenotype, because we delineated the importance of constitutive Trp degradation to sustain the malignant phenotype of cancer by acting on the tumour cells themselves.

Emerging evidence indicates a tumour-promoting role of the AHR. AHR activation promotes clonogenicity and invasiveness of cancer cells^{11,17}. Transgenic mice with a constitutively active AHR spontaneously develop tumours^{18,19}, and the repressor of the AHR (AHRR) represents a tumour suppressor in multiple human cancers²⁰. The aberrant phenotype of *Ahr*-deficient mice points to the existence of endogenous AHR ligands²¹. Although different endogenously produced metabolites such as arachidonic acid metabolites, bilirubin, cyclic AMP, tryptamine and 6-formylindolo[3,2-*b*]carbazole (FICZ) have been shown to be agonists of the AHR²², their functionality has not been convincingly demonstrated in a pathophysiological context such as cancer or immune activation. The search for endogenous ligands of the AHR is therefore continuing.

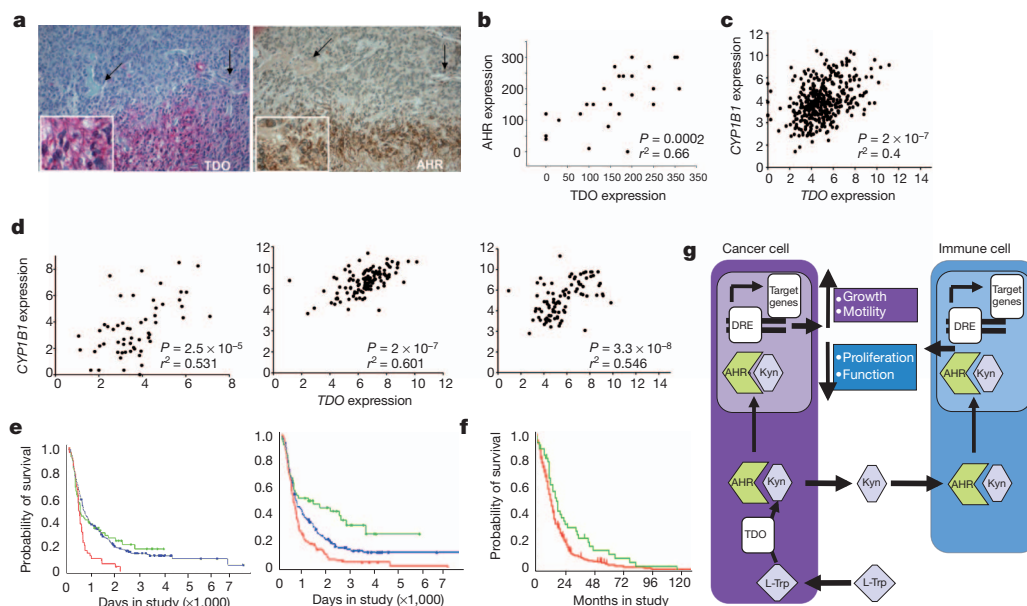


Figure 5 | TDO-derived Kyn activates the AHR in diverse human cancers, and AHR activation predicts survival in patients with glioma. **a**, Correlation of TDO expression (red) and AHR expression (brown) in consecutive sections of human glioblastoma tissue. Arrows indicate vessels for orientation. Magnification $\times 40$; insets $\times 200$. **b**, Correlation between TDO and AHR expression in human glioma tissue based on H-scores of TDO and AHR, calculated by Spearman rank correlation ($n = 26$). **c**, Correlation between TDO and *CYP1B1* expression in microarray data of human glioblastoma ($n = 396$) analysed by Spearman rank correlation. **d**, Correlation between TDO and *CYP1B1* expression in microarray data of human bladder cancer (left; $n = 58$), human lung cancer (centre; $n = 122$) and human ovarian carcinoma (right; $n = 91$) analysed by Spearman rank correlation. **e**, Survival probabilities of patients with glioma (WHO grades II–IV) with high expression (red) of TDO (left) or AHR (right) compared with those of patients with intermediate (blue) or low (green) expression of these genes derived from Rembrandt¹³. **f**, Survival probabilities of patients with glioblastoma with high expression (red) of the AHR target gene *CYP1B1* compared with those of patients with low (green) expression of *CYP1B1* derived from the glioblastoma data set of the Cancer Genome Atlas network¹⁴ ($n = 362$). **g**, Synoptic figure highlighting the autocrine and paracrine effects of TDO-derived Kyn on cancer cells and immune cells through the AHR.

g, Synoptic figure highlighting the autocrine and paracrine effects of TDO-derived Kyn on cancer cells and immune cells through the AHR.

We now link these two important pathways contributing to cancer progression by showing that Trp catabolism leads to AHR activation, and we provide evidence of a pathophysiological human condition that is associated with the production of sufficient amounts of a functionally relevant endogenous AHR ligand. Our results reveal a differential response of primary immune cells and transformed cancer cells to AHR-mediated signals, which is in line with various toxicological studies using the classical exogenous AHR ligands TCDD and 3-methylcholanthrene^{11,17,23}. Exposure to these xenobiotics leads to profound suppression of cellular and humoral immune responses²³, while also promoting carcinogenesis and inducing tumour growth^{11,17}. These cell-specific differences in AHR effects are likely to depend on the expression of factors differentially regulating AHR signal transduction such as the AHRR²⁰ as well as cell-specific transcription factor crosstalk shaping the response to AHR activation²⁴.

It is likely that Kyn-mediated activation of the AHR is not only relevant in the setting of cancer. For instance, activation of the mouse and human AHR by agonistic ligands induces regulatory T cells^{25–28}. Mice with a poor-affinity AHR suffer from exacerbated autoimmune encephalomyelitis in the absence of an exogenous ligand²⁶, and Trp catabolites suppress autoimmune neuroinflammation^{29,30}, suggesting that activation of Trp catabolism represents an endogenous feedback loop to restrict inflammation through the AHR. In fact, exogenous Kyn is involved in the regulation of immune cells in mice through the AHR^{31,32}. Kyn concentrations sufficient to activate the AHR are also generated by IDO in response to inflammatory stimuli (Supplementary Fig. 11a–c). In a broader context, a significant number of malignancies arise from areas of mostly chronic infection and inflammation³³, where Trp catabolism in the tumour microenvironment is activated and sustains local immune suppression³⁴. Activation of the AHR by Kyn generated in response to inflammatory stimuli may thus constitute a previously unrecognized pathway connecting inflammation and carcinogenesis.

METHODS SUMMARY

TDO expression was analysed by immunohistochemistry in human tumours. Its relevance for Trp degradation was determined by genetic knockdown or over-expression of TDO. Trp and Kyn were measured in cell culture supernatants, human sera and xenograft tissue by high-performance liquid chromatography. Mixed leukocyte reactions, chromium release, ELISpot and staining of immune cells in tumour tissues were used to assess the immune effects of TDO activity. Cell cycle analysis, Matrigel and spheroid invasion assays, scratch assays, sphere formation assays and clonogenicity assays were employed to analyse the autocrine effects of TDO activity. All animal procedures followed the institutional laboratory animal research guidelines and were approved by the governmental authorities. Orthotopic implantation of human glioma cells with and without stable knockdown of TDO into *CD1nu/nu* mice, subcutaneous injection of these cells into NK-depleted or wild-type *CD1nu/nu* mice and subcutaneous injection of murine *Tdo*-proficient and *Tdo*-deficient GL261 cells into syngeneic C57BL/6N mice were performed to analyse the autocrine and paracrine effects of TDO activity *in vivo*. Microarray analysis of Kyn-treated human glioma cells was performed to identify signalling pathways activated by Kyn. Analysis of AHR translocation, DRE-luciferase assays and radioligand binding assays confirmed activation of the AHR by Kyn. Pharmacological inhibition and stable knockdown of the AHR (*in vitro* and *in vivo*) proved that the effects of Kyn are AHR dependent. Injection of *Tdo*-proficient and *Tdo*-deficient tumour cells into *Ahr*^{+/+} and *Ahr*^{-/-} mice was used to address the contribution of host effects to TDO-mediated cancer promotion. Finally, stainings, microarray data and clinical data on human tumour tissues were used to analyse whether TDO activates the AHR in human cancers and how this affects survival.

Received 17 November 2010; accepted 17 August 2011.

Published online 5 October 2011.

1. Muller, A. J. *et al.* Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene *Bin1*, potentiates cancer chemotherapy. *Nature Med.* **11**, 312–319 (2005).
2. Munn, D. H. & Mellor, A. L. Indoleamine 2,3-dioxygenase and tumor-induced tolerance. *J. Clin. Invest.* **117**, 1147–1154 (2007).

3. Uyttenhove, C. *et al.* Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase. *Nature Med.* **9**, 1269–1274 (2003).
4. Ball, H. J. *et al.* Characterization of an indoleamine 2,3-dioxygenase-like protein found in humans and mice. *Gene* **396**, 203–213 (2007).
5. Metz, R. *et al.* Novel tryptophan catabolic enzyme IDO2 is the preferred biochemical target of the antitumor indoleamine 2,3-dioxygenase inhibitory compound D-1-methyl-tryptophan. *Cancer Res.* **67**, 7082–7087 (2007).
6. Lob, S. *et al.* Inhibitors of indoleamine-2,3-dioxygenase for cancer therapy: can we see the wood for the trees? *Nature Rev. Cancer* **9**, 445–452 (2009).
7. NewLink Genetics Corporation. IDO Inhibitor Study for Relapsed or Refractory Solid Tumors (D-1MT) <<http://clinicaltrials.gov/show/NCT00739609>> (2008).
8. Miller, C. L. *et al.* Expression of the kynurenine pathway enzyme tryptophan 2,3-dioxygenase is increased in the frontal cortex of individuals with schizophrenia. *Neurobiol. Dis.* **15**, 618–629 (2004).
9. Frumento, G. *et al.* Tryptophan-derived catabolites are responsible for inhibition of T and natural killer cell proliferation induced by indoleamine 2,3-dioxygenase. *J. Exp. Med.* **196**, 459–468 (2002).
10. Denison, M. S. & Nagy, S. R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334 (2003).
11. Gramatzki, D. *et al.* Aryl hydrocarbon receptor inhibition downregulates the TGF- β /Smad pathway in human glioblastoma cells. *Oncogene* **28**, 2593–2605 (2009).
12. Reyes, H., Reisz-Porszasz, S. & Hankinson, O. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193–1195 (1992).
13. National Cancer Institute. REMBRANDT Home Page (<http://rembrandt.nci.nih.gov>) (2005).
14. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
15. Opitz, C. A. *et al.* The indoleamine 2,3-dioxygenase (IDO) inhibitor 1-methyl-D-tryptophan upregulates IDO1 in human cancer cells. *PLoS ONE* **6**, e19823 (2011).
16. Dolusic, E. *et al.* Tryptophan 2,3-dioxygenase (TDO) inhibitors. 3-(2-(pyridyl)ethenyl)indoles as potential anticancer immunomodulators. *J. Med. Chem.* **54**, 5320–5334 (2011).
17. Bui, L. C. *et al.* Nedd9/Hef1/Cas-L mediates the effects of environmental pollutants on cell migration and plasticity. *Oncogene* **28**, 3642–3651 (2009).
18. Andersson, P. *et al.* A constitutively active dioxin/aryl hydrocarbon receptor induces stomach tumors. *Proc. Natl Acad. Sci. USA* **99**, 9990–9995 (2002).
19. Moennikes, O. *et al.* A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice. *Cancer Res.* **64**, 4707–4710 (2004).
20. Zudaire, E. *et al.* The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. *J. Clin. Invest.* **118**, 640–650 (2008).
21. Fernandez-Salguero, P. *et al.* Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor. *Science* **268**, 722–726 (1995).
22. Nguyen, L. P. & Bradfield, C. A. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem. Res. Toxicol.* **21**, 102–116 (2008).
23. Esser, C., Rannug, A. & Stockinger, B. The aryl hydrocarbon receptor in immunity. *Trends Immunol.* **30**, 447–454 (2009).
24. Frericks, M., Burgoon, L. D., Zacharewski, T. R. & Esser, C. Promoter analysis of TCDD-inducible genes in a thymic epithelial cell line indicates the potential for cell-specific transcription factor crosstalk in the AhR response. *Toxicol. Appl. Pharmacol.* **232**, 268–279 (2008).
25. Apetoh, L. *et al.* The aryl hydrocarbon receptor interacts with c-Maf to promote the differentiation of type 1 regulatory T cells induced by IL-27. *Nature Immunol.* **11**, 854–861 (2010).
26. Quintana, F. J. *et al.* Control of T_H17 and T_H17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65–71 (2008).
27. Quintana, F. J. *et al.* An endogenous aryl hydrocarbon receptor ligand acts on dendritic cells and T cells to suppress experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **107**, 20768–20773 (2010).
28. Veldhoen, M. *et al.* The aryl hydrocarbon receptor links T_H17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–109 (2008).
29. Platten, M. *et al.* Treatment of autoimmune neuroinflammation with a synthetic tryptophan metabolite. *Science* **310**, 850–855 (2005).
30. Opitz, C. A., Wick, W., Steinman, L. & Platten, M. Tryptophan degradation in autoimmune diseases. *Cell. Mol. Life Sci.* **64**, 2542–2563 (2007).
31. Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *J. Immunol.* **185**, 3190–3198 (2010).
32. Nguyen, N. T. *et al.* Aryl hydrocarbon receptor negatively regulates dendritic cell immunogenicity via a kynurenine-dependent mechanism. *Proc. Natl Acad. Sci. USA* **107**, 19961–19966 (2010).
33. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
34. Muller, A. J. *et al.* Chronic inflammation that facilitates tumor progression creates local immune suppression by inducing indoleamine 2,3 dioxygenase. *Proc. Natl Acad. Sci. USA* **105**, 17073–17078 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Rauschenbach, J. Reiart, S. Koch and A. Mlitzko for technical assistance, P.-N. Pfenning for help with the invasion assays, J. Blaes for help generating genetically modified cells, T. V. Lanz and I. Oezgen for help with animal experiments, C. Esser for providing *Ahr*^{-/-} mice, M. Schwarz for providing the human DRE-luciferase construct, D. Lemke for generating GICs, K. Ochs for providing human

umbilical-vein endothelial cell cDNA, A. Hertenstein for generation of CD4⁺ and CD8⁺ T cells, M. Betts and R. B. Russel for modelling of the binding of Kyn to the AHR, R. Koch for performing quantitative reverse transcriptase polymerase chain reaction analyses, R. Tudoran for generation of GL261 cells overexpressing murine TDO, M. Deponte for help with the radioligand binding assay, D. Schemmer for collecting and banking serum samples, W. Roth for providing tissue specimens, M. Remke for suggestions regarding data analysis, and G. Hämmerling, B. Arnold and M. Feuerer for discussions. This work was supported by grants from the Helmholtz Association (VH-NG-306) to M.P., the German Research Foundation to M.P. and W.W. (Deutsche Forschungsgemeinschaft SFB 938 TP K), the Hertie Foundation to W.W. and the Helmholtz Alliance on Systems Biology to S.T. and I.L. T.S. is supported by a DKFZ PhD Program stipend. C.A.O. is supported by a Heidelberg University Medical Faculty Postdoctoral Fellowship.

Author Contributions C.A.O. and U.M.L. contributed equally to this study, designed and performed experiments, analysed data and wrote the paper. F.S. and A.D. analysed

protein expression by immunohistochemistry. I.T. cloned constructs and designed experiments. S.T. and I.L. performed nuclear translocation assays. M.O. performed animal experiments. T.S. performed immune experiments. L.J. and M.J. performed MRI scans. C.L.M. and G.J.G. provided antibodies and designed experiments. D.S. performed and analysed EROD and DRE–luciferase assays. C.L. synthesized the TDO inhibitor. M.W. and W.W. were involved in study design and data interpretation. B.R. analysed microarray data. M.P. and C.A.O. conceptualized the study, interpreted data, designed experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Microarray data were deposited in the Gene Expression Omnibus repository (GEO) at www.ncbi.nlm.nih.gov/geo/ under accession number GSE25272. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.P. (m.platten@dkfz-heidelberg.de).

Diffraction-unlimited all-optical imaging and writing with a photochromic GFP

Tim Grotjohann^{1*}, Ilaria Testa^{1*}, Marcel Leutenegger^{1*}, Hannes Bock¹, Nicolai T. Urban¹, Flavie Lavoie-Cardinal¹, Katrin I. Willig¹, Christian Eggeling¹, Stefan Jakobs^{1,2} & Stefan W. Hell¹

Lens-based optical microscopy failed to discern fluorescent features closer than 200 nm for decades, but the recent breaking of the diffraction resolution barrier by sequentially switching the fluorescence capability of adjacent features on and off is making nanoscale imaging routine. Reported fluorescence nanoscopy variants switch these features either with intense beams at defined positions or randomly, molecule by molecule. Here we demonstrate an optical nanoscopy that records raw data images from living cells and tissues with low levels of light. This advance has been facilitated by the generation of reversibly switchable enhanced green fluorescent protein (rsEGFP), a fluorescent protein that can be reversibly photoswitched more than a thousand times. Distributions of functional rsEGFP-fusion proteins in living bacteria and mammalian cells are imaged at <40-nanometre resolution. Dendritic spines in living brain slices are super-resolved with about a million times lower light intensities than before. The reversible switching also enables all-optical writing of features with subdiffraction size and spacings, which can be used for data storage.

In a fluorescence microscope, diffraction prevents (excitation) light being focused more sharply than $\lambda/(2\text{NA})$, with λ being the wavelength of light and NA the numerical aperture of the lens. Thus, as they are illuminated together, features residing any closer together than this distance also fluoresce together and appear in the image as a single blur. The diffraction resolution barrier can be overcome by forcing such nearby features to fluoresce sequentially, but this strategy clearly requires a mechanism for keeping fluorophores that are exposed to excitation light non-fluorescent^{1–3}.

In stimulated emission depletion (STED) microscopy^{1,4}, this is accomplished by the so-called STED beam, which turns the fluorescence capability of fluorophores off by a photon-induced de-excitation. Because at least a single de-exciting photon must be available within the lifetime ($\tau \approx 1\text{--}5\text{ ns}$) of the fluorescent molecular state, the intensity of the focal STED beam must exceed the threshold $I_s = C\tau^{-1}$ with C accounting for the probability of a STED beam photon to interact with the fluorophore^{1,4}. The STED beam, usually formed as a doughnut overlaid with the excitation beam, features a central point of zero intensity at which the fluorophores can still assume the fluorescent state. As this point can be positioned with arbitrary precision in space, the coordinate of the emitting (on-state) fluorophores is known at any instant: it is the position of zero intensity^{3,5,6} and its immediate vicinity, where the STED beam is still weaker than I_s . The diameter of this area is given by $d \approx \lambda/[2\text{NA} \times (1 + I_m/I_s)^{1/2}]$, with I_m (typically $\gg I_s$) denoting the intensity at the doughnut crest. Hence, features that are (just slightly) more apart than $d \ll \lambda/(2\text{NA})$ cannot fluoresce at the same time even when simultaneously illuminated by excitation light⁶. Scanning the beams across the sample and recording the fluorescence yields images of subdiffraction resolution d automatically and irrespective of the fluorophore concentration in the sample.

De-excitation by stimulated emission is the most basic and general mechanism for modulating the fluorescence ability of a molecule. However, by requiring light intensities $> I_s \approx 1\text{--}10\text{ MW cm}^{-2}$, attaining high resolutions by this mechanism necessitates large I_m values. For

example, $d < 40\text{ nm}$ typically entails $I_m = 100\text{--}500\text{ MW cm}^{-2}$ (ref. 6). Although intensities of this order have been demonstrated to be live-cell compatible^{4,7–10}, all-optical nanoscopy methods operating at fundamentally lower light levels are highly in demand^{2,5,11–13}, because they allow larger fields of view^{5,14} and can avoid photodamage.

A route to low light level operation is to replace STED with a fluorescence switching mechanism having a lower threshold I_s (refs 2, 5, 11–13). Following the equation for I_s , this can be realized by exploiting transitions between fluorophore states of longer lifetime $\tau \gg 1\text{ }\mu\text{s}$ (refs 2, 5, 11). Hence, it has been suggested that fluorescence can be switched by transferring the fluorophores transiently to a generic metastable dark (triplet) state of $\tau \approx 10^{-3}\text{--}100\text{ ms}$ (refs 2, 15). A more attractive option is to use fluorophores that can be explicitly 'photoswitched'^{5,11}, for example, by photoisomerization. Hence, in 2003 it was proposed to implement a STED-like microscope with STED being replaced by a reversible on-off switch as encountered in organic photochromic fluorophores and reversibly photoswitchable fluorescent proteins (RSFPs)^{5,11}.

In fact, this strategy is even more general because any reversible transition between a signalling and a non-signalling state can be used for breaking the diffraction barrier¹⁵. Therefore, all concepts that switch the fluorescence capability of molecules at sample coordinates predefined by patterns of light have been generalized under the name RESOLFT^{6,12}, which stands for reversible saturable optical (fluorescence) transition between two states. Note that a photoswitch is a perfect saturable transition. Concomitantly, the concept was extended to subdiffraction writing^{11,13} and data storage, in which case the on-state is a reactive state from which the molecule can be made permanent whereas the off-state serves as a temporary 'mask' defining the structure to be written.

Super-resolution by switching RSFPs was shown in 2005¹⁶, but this study relied on aFP595¹⁷, a tetrameric protein with low fluorescence quantum yield. Moreover, when translating the light pattern across the sample, the proteins faded after a few cycles, implying that features that had been turned off could not be turned on again in order to be

¹Department of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. ²University of Göttingen Medical School, Robert-Koch-Str. 40, 37075 Göttingen, Germany.

*These authors contributed equally to this work.

read out. Biological imaging therefore remained unviable¹⁸. Other studies using a variant of the RSFP called dronpa¹⁹ faced the same challenge²⁰. As a rule of thumb, an m -fold resolution improvement along a certain direction requires $\sim m$ switching cycles, meaning that $m = 10$ along the x - and y -axes entails $\sim m^2 = 100$ cycles, whereas $\sim 1,000$ cycles are required for x , y and z (ref. 6). Thus, for RESOLFT super-resolution, the number of switching cycles afforded by the fluorophore assumes a vital role.

Because they are able to generate an image with a single on-off cycle^{3,6,21}, the super-resolution concepts called (F)PALM^{22,23} and STORM^{21,24}, which have emerged in the interim, have successfully harnessed the switching between metastable states for gaining subdiffraction resolution. However, these methods rely on the imaging and computation-aided localization of individual fluorophores amidst the scattering and autofluorescence background common in (living) cells and tissues. Moreover, rapid localization of a sufficiently large number of fluorophores requires the excitation light to be intense^{22,23}. In contrast, a RESOLFT approach is able to instantly record the emission from all fluorophores attached to the nanosized feature of interest⁶, and can be easily combined with confocal microscopy for three-dimensional imaging and background suppression. Yet again, because all RSFPs, conventional fluorescent proteins²⁵ and photochromic rhodamines²⁶ seemed unsuitable (Supplementary Fig. 1), an all-optical nanoscopy approach operating at low light levels appeared unviable.

Similarly, although STED/RESOLFT-inspired optical writing with photochromic compounds has been shown to yield structures $< \lambda / (2\text{NA})$, writing such structures with spacings $< \lambda / (2\text{NA})$ remained challenging^{27–30}, again the impediment being the requirement of many on-off cycles before the structure is made permanent. Here we introduce a RSFP enabling both low-light-level all-optical nanoscopy of living cells and tissues, and far-field optical writing and reading of patterns of subdiffraction size and density.

Generating a reversibly switchable GFP

All fluorescent proteins have a similar fold, namely an 11-stranded β -barrel with a central helix containing the chromophore, which is typically in a *cis*-configuration³¹. Light-driven switching of RSFPs generally involves an isomerization of the chromophore, frequently coupled with a change of its protonation state^{32–36}. We started from EGFP³⁷ and identified, using its X-ray structure³¹, amino acid residues the exchange of which was expected to facilitate isomerization. We expressed numerous EGFP variants in *Escherichia coli* and screened for colonies expressing an RSFP with an automated microscope. To this end, we alternated site-directed and error-prone mutagenesis while maintaining the key amino acids of EGFP (that is, F64L and S65T)³⁷; we concomitantly introduced A206K to ensure that the protein remained a monomer³⁸.

The amino acid exchange Q69L was sufficient to make EGFP(A206K) reversibly switchable, but the resulting on-off contrast was low. Although it makes the protein switchable³⁹, we avoided the mutation E222Q because it seemed to reduce the number of cycles. After analysing $\sim 30,000$ clones, we identified EGFP(Q69L/V150A/V163S/S205N/A206K) (Supplementary Fig. 2) that could be reversibly switched on at $\lambda = 405$ nm and off at 491 nm, and named it reversibly switchable EGFP (rsEGFP).

At equilibrium, rsEGFP adopts a bright on-state (fluorescence quantum yield $\Phi_{\text{FL}} = 0.36$; extinction coefficient $\varepsilon = 47,000 \text{ M}^{-1} \text{ cm}^{-1}$ (Supplementary Table 1)). In the on-state, rsEGFP exhibits a single absorption band peaking at 491 nm (Fig. 1a), corresponding to the ionized state of the phenolic hydroxyl of the chromophore⁴⁰. The pK_{a} of the chromophore is 6.5 (Supplementary Fig. 3). Absorption at 490 nm yields fluorescence peaking at 510 nm and, in a competing process, switches rsEGFP off (Figs 1a–c). Prolonged irradiation of a pH 7.5 solution of purified rsEGFP at ~ 490 nm reduces the rsEGFP fluorescence to 1–2% of its initial value. The off-state exhibits a single absorption band

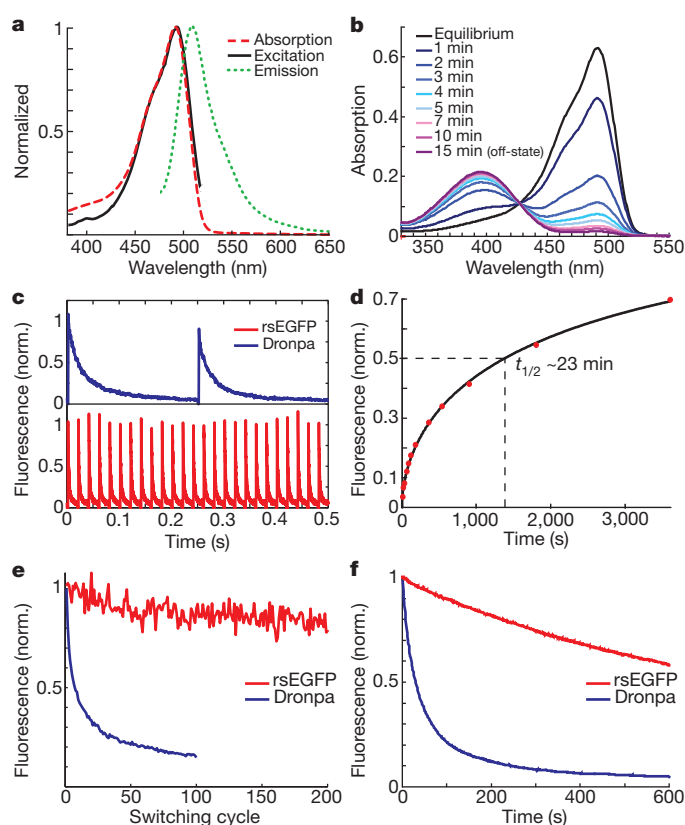


Figure 1 | Properties of rsEGFP. **a**, Absorption (red dashed line), excitation (solid black line) and fluorescence (dotted green line) spectrum of rsEGFP in the fluorescent equilibrium state at pH 7.5. **b**, Absorption spectra obtained at different time points during irradiation with 488-nm light. **c**, Switching curves of dronpa (blue) and rsEGFP (red) immobilized in PAA using the same intensities. Switching was performed by alternating irradiation at 405 nm (2 kW cm^{-2}) and at 491 nm (0.6 kW cm^{-2}). The duration of off-switching at 491 nm was chosen such that the fluorescence reached a minimum; irradiation with 405 nm was chosen so that the proteins were fully switched. **d**, Relaxation of rsEGFP embedded in PAA from the off-state into the fluorescent equilibrium state at 22°C . The black line is a stretched exponential fit with a stretching factor of ~ 0.6 accounting for inhomogeneous spectral broadening or the involvement of multiple dark states. **e**, Fluorescence per switching cycle normalized to the initial fluorescence, with the same light intensities and switching durations as in **c**. **f**, Photobleaching: rsEGFP and dronpa embedded in a PAA layer were kept in their on-states by continuous irradiation at 405 nm (1 kW cm^{-2}), while fluorescence was probed by irradiation at 491 nm (3 kW cm^{-2}).

at 396 nm, corresponding to the neutral state of the chromophore (Fig. 1b). Excitation at this band switches the protein back to the on-state. At room temperature rsEGFP converts spontaneously from the off- into the on-state with a half-time of ~ 23 min (Fig. 1d).

We compared the properties of rsEGFP with that of the well-known RSFP dronpa¹⁹. With the proteins embedded in a 12.5% polyacrylamide (PAA) layer and using light of 491 nm (0.6 kW cm^{-2}) and 405 nm (2 kW cm^{-2}), a complete on-off cycle took 250 ms for dronpa and 20 ms for rsEGFP (Fig. 1c). Dronpa went through < 10 cycles before its fluorescence was reduced to 50%, whereas rsEGFP went through $\sim 1,200$ cycles under the same conditions (Fig. 1e). To compare bleaching, dronpa and rsEGFP were kept in the on-state by continuous irradiation at 405 nm (1 kW cm^{-2}) while fluorescence was generated by irradiation at 491 nm (3 kW cm^{-2}). Whereas dronpa fluorescence was reduced to 50% within $t_{1/2} \approx 30$ s, for rsEGFP we measured $t_{1/2} \approx 800$ s (Fig. 1f). The rsEGFP chromophore matured with a half-time of ~ 3 h at 37°C (Supplementary Fig. 4). The protein behaved as a monomer *in vitro* (Supplementary Fig. 5), could be fused to various proteins, including α -tubulin and histone H2B (Supplementary Fig. 6), and was repeatedly switchable in living cells (Supplementary Fig. 7).

Rewritable data storage

To analyse whether immobilized rsEGFP could be used for repeated short-term data storage⁴¹, we coated a microscope slide with a $<1\text{-}\mu\text{m}$ thin layer of rsEGFP ($\sim 0.03\text{ mM}$) in PAA. Switching and reading by illumination at 405 nm and 491 nm in a scanning confocal set-up provided an on-off contrast of $\sim 50:1$. We translated the text of 25 Grimm's fairy stories (<http://www.gutenberg.org/files/11027/11027.txt>) into 7-bit binary ASCII code ('0': off; '1': on) and wrote and read the $\sim 270,000$ letters into a $17\text{ }\mu\text{m} \times 17\text{ }\mu\text{m}$ region in 6,596 frames, each comprising 41 letters (287 bits) (Fig. 2). Individual bits were $\sim 0.5\text{ }\mu\text{m}$ in diameter with $1\text{ }\mu\text{m}$ centre-to-centre spacing, corresponding to a DVD storage density. Discriminating '0' from '1' by a simple threshold entailed 7 bit errors within the entire data set. After $\sim 6,600$ read/write cycles in the same region, the average fluorescence of the '1' was reduced by $\sim 35\%$ (Supplementary Fig. 8). Hence, the same rsEGFP layer can be used for $\sim 15,000$ read/write processes.

RESOLFT nanoscopy of living samples

Next, we implemented a scanning confocal set-up with a 405 nm (ultraviolet) beam for switching the rsEGFP on, a 491 nm (blue) beam for eliciting fluorescence, and a doughnut-shaped 491 nm beam for the off-switching (Supplementary Fig. 9). We fused rsEGFP to the amino-terminus of the bacterial actin homologue MreB⁴² and expressed the fusion protein in *E. coli* bacteria. Living bacteria on agar-coated slides were recorded by first irradiating each pixel for $100\text{ }\mu\text{s}$ with ultraviolet light (1 kW cm^{-2}), thus activating most of the rsEGFP in the focal volume. Then the doughnut-shaped blue beam ($I_m \approx 1\text{ kW cm}^{-2}$) was applied for 10–20 ms to switch all the rsEGFP molecules off, except those located within $d/2$ distance from the doughnut centre. Lastly, the rsEGFP fluorescence was read out for 1–2 ms by the 491-nm beam ($\sim 1\text{ kW cm}^{-2}$). The sequence was repeated for each sample pixel.

The double-helical cytoskeletal structure of rsEGFP–MreB is more clearly revealed by RESOLFT than by its confocal counterpart

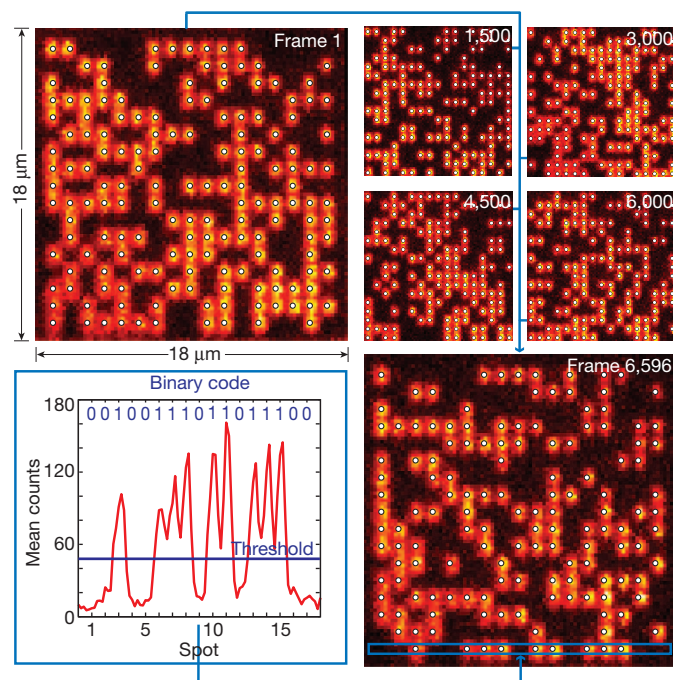


Figure 2 | Rewritable data storage. The text of 25 Grimm's fairy stories (ASCII code; 1.9 Mbits) consecutively written and read on a $17 \times 17\text{ }\mu\text{m}$ area of a PAA layer containing rsEGFP, with bits written as spots (representative frames shown). The white dots mark spots that were recognized as set bits ('1's). The graph shows an intensity profile along the indicated area, averaged over three pixels along the y-axis. The blue line indicates the threshold used to assign read spots to '0's or '1's'.

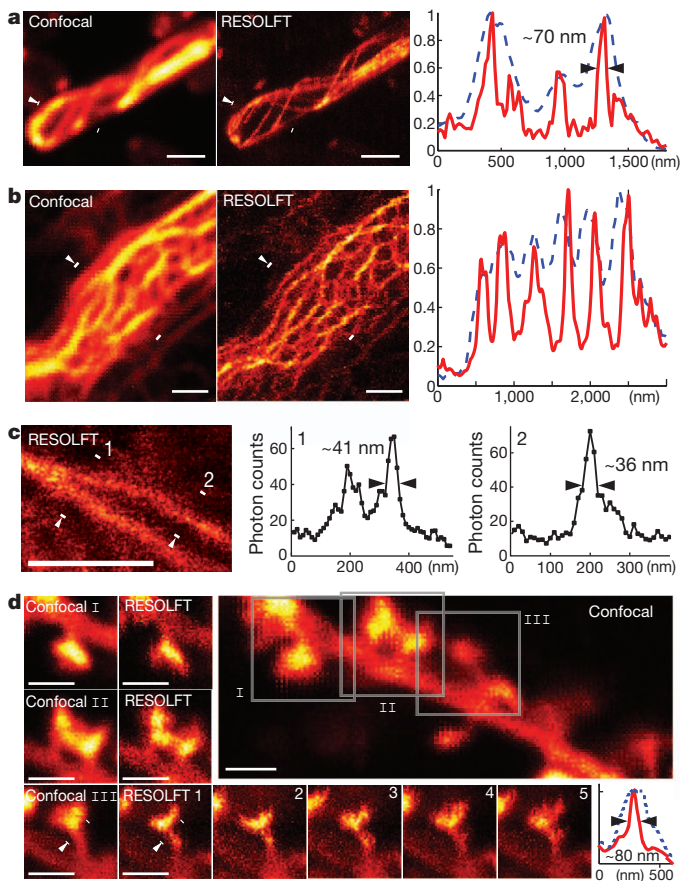


Figure 3 | RESOLFT nanoscopy of living cells. **a**, *E. coli* bacterium expressing rsEGFP–MreB: confocal (left) and corresponding RESOLFT (middle) image. **b**, Mammalian (PtK2) cell expressing keratin-19–rsEGFP imaged in the confocal (left) and the RESOLFT (middle) mode. **a**, **b**, Graphs show the normalized fluorescence profiles between the two white markers with the white arrowhead indicating the direction (solid red, RESOLFT; dashed blue, confocal). **c**, RESOLFT image (left) of keratin-19–rsEGFP filaments in a PtK2 cell recorded with a pixel size of $10\text{ nm} \times 10\text{ nm}$; smoothed with a low-pass Gaussian filter of 1.2 pixel width. Graphs 1 and 2 extracted from the image as indicated reveal resolution $d < 40\text{ nm}$. **d**, Dendrite within a living organotypic hippocampal slice expressing lifeact–rsEGFP. Main image: confocal overview. I–III: three spines, as indicated on the main image, each imaged in the confocal (left) and the RESOLFT mode (right). Spine III was repeatedly imaged in the RESOLFT mode within 5 min, demonstrating the changes over time. Graph: normalized profile across a spine neck as imaged in the RESOLFT (solid red) or the confocal mode (dashed blue) between the two white markers. Scale bars, $1\text{ }\mu\text{m}$.

(Fig. 3a). The RESOLFT image of a typical filament showed a full-width half-maximum (FWHM) of $\sim 70\text{ nm}$. Because this value seemed to be determined by the thickness of the filament itself, a more accurate upper limit for the resolution d is obtained by imaging the finer keratin-19–rsEGFP intermediate filament network in living mammalian cells (Fig. 3b, c). Line profiles from recorded data gave $d < 40\text{ nm}$ corresponding to a 5–6-fold all-optical resolution improvement over confocal microscopy (Fig. 3c).

To investigate its applicability to living brain tissue, we locally injected viral particles carrying a lifeact–rsEGFP construct into a cultured organotypic hippocampal brain slice. Lifeact is a 17-amino-acid-long peptide with high affinity to filamentous actin⁴³. RESOLFT revealed fine morphological differences between the spines protruding from a dendrite (Fig. 3d). A profile through a spine neck showed a FWHM of $< 80\text{ nm}$. Electron microscopy of similar samples demonstrated that this value is close to the actual size of the spine necks themselves⁴⁴, suggesting a resolution d substantially $< 80\text{ nm}$. Repeated imaging revealed dynamic changes over 5 min (Fig. 3d). Altogether, the resolution is comparable to that provided by STED

on similar structures¹⁰, but here it is obtained with light intensities lower by about a million times.

RESOLFT optical data storage

For investigating subdiffraction resolution writing, an rsEGFP layer was prepared as previously outlined. The writing entailed (1) an ultraviolet beam (405 nm, 1 kW cm^{-2}) applied for 100 μs to switch rsEGFP on, (2) a 2-ms break for equilibration, (3) a doughnut-shaped blue beam (491 nm, 0.5 kW cm^{-2}) lasting 20 ms confining the on-state within $d/2$ around the doughnut centre, and (4) an ~ 2 -ms 532 nm beam ($\sim 900 \text{ kW cm}^{-2}$) for transferring on-state rsEGFP to a permanent off (bleached) state (Fig. 4a) (Supplementary Fig. 10a). Lastly, the rsEGFP molecules located outside this region were switched back on, which is critical for writing another feature within subdiffraction proximity.

We wrote nine patterns of 3×3 bit fields in an rsEGFP layer, with 250 nm centre-to-centre separation between individual bits (Fig. 4b), both in the conventional and in the RESOLFT mode. Whereas conventional writing and/or confocal reading blurred the data, the bits were fully discernible when both writing and reading were performed by RESOLFT. We wrote and read the data down to distances of 200 nm between the individual bits (Supplementary Fig. 10b). Hence this scheme allowed storing and reading out bits ~ 4 times more densely than by regular focusing. The structures could be read 5–10 times.

Discussion and conclusion

The many-switching cycles afforded by the fluorescence protein rsEGFP reported here has facilitated live-cell RESOLFT microscopy, a super-resolution microscopy that is similar to STED microscopy in usability but operates at $\sim 10^6$ times lower levels of light. Multiphoton-induced optical damage⁴⁵ can therefore be virtually excluded. The fundamental reduction in optical intensity required for the on-off

switching stems from the fact that the fluorescence capability of the molecule is not modulated by disallowing the population of its nano-second fluorescent state, but rather by toggling it between two long-lived ground states, one in which the fluorophore remains dark when exposed to the excitation light.

RESOLFT is readily combined with confocal imaging, which increases its use in scattering living samples. In fact, the imaging of neuronal spines in living organotypical brain slices testifies this potential. Although the recording time reported here is still of the order of most other super-resolution techniques^{3,6,21} and slower than the fastest biological STED recordings⁷, by gathering the signal from typically many molecules located at predefined positions, RESOLFT has all the prerequisites for fast imaging. Scanning with arrays of doughnuts or zero-intensity lines (so-called structured illumination^{5,14,18,46}) and detection by a camera will reduce the number of scanning steps required to cover large fields of view and facilitate low-intensity video-rate imaging. The maximum recording speed is determined by the time it takes to establish the disparity of (on-off) states in space, that is, by the switching kinetics, which probably can be improved by further mutagenesis. Note that the switching is not restricted to changes in brightness (on-off) only. Other reversible transitions between disparate states may also prove suitable for RESOLFT imaging, such as states yielding differences in emission wavelengths, lifetime or polarization.

Photoswitching between long-lived states also poses challenges, because in the process the molecule can assume transient (dark) states, such as triplet states, which depend on the molecular micro-environment. In this regard, STED maintains a unique advantage because it entails just basic optical transitions between the ground and the fluorescent state; no atom relocation, spin flip or change in chemical bond is required to switch the fluorescence capability of the molecule — just light. Therefore, switching fluorescence by STED is nearly universal and instantaneous.

The switching stamina of rsEGFP also enabled writing and reading of patterns of both subdiffraction size and spacing d , which has so far been difficult for direct far-field optical writing. In our study, the smallest obtainable structure size was co-determined by the fact that the 532-nm light moderately bleached the off-state proteins too, thus reducing the writing contrast. However, this initial demonstration should spur on new advancements in this field, because current nanowriting efforts are dominated by concepts that resort to much shorter wavelengths of electromagnetic radiation at which focusing becomes exceedingly difficult. In fact, RESOLFT and related concepts are unique for creating materials that are nanostructured in three dimensions³⁰. To maximize the resolution along the optical axis (z), RESOLFT imaging and writing can also be combined with 4Pi microscopy⁴⁷, in which case three-dimensional resolution of <10 nm should become possible at ultralow light levels.

The resolution demonstrated here is similar or even exceeds the resolution attained until now by STED in living cells^{8,10}. Although in both methods the resolution can be continually increased by increasing I_m/I_s , in STED microscopy this strategy will reach practical limits due to the intensities required. Using a threshold intensity I_s that is lower by many orders of magnitude, switching between long-lived states overcomes these limits and, as we have demonstrated here, offers a pathway to lens-based optical imaging and writing at molecular dimensions.

METHODS SUMMARY

Protein generation and screening. Site-directed mutagenesis was performed with the QuikChange Site Directed Mutagenesis Kit (Stratagene) or a multiple-site approach using several degenerative primers. The proteins were expressed from the high-copy expression vector pQE31 (Qiagen) and expressed in *E. coli*.

Viral transfection. A modified Semliki Forest Virus containing the pSCA-Lifeact-rsEGFP vector construct was injected into the slice cultures using a patch pipette. Imaging was performed within 16–48 h after incubation.

Data storage. A layer containing immobilized rsEGFP was prepared by mixing 24.5 μl purified proteins (0.09 mM) with 17.5 μl Tris-HCl pH 7.5, 30 μl acrylamide (Rotiphorese Gel 30, Roth), 0.75 μl 10% ammonium persulfate and 1 μl 10%

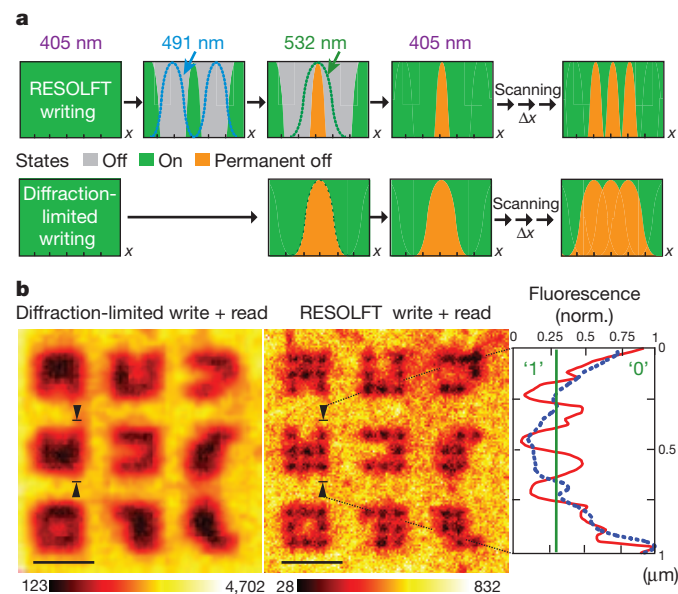


Figure 4 | Subdiffraction-resolution writing and reading using rsEGFP and visible light. **a**, Top, schematic of RESOLFT writing: rsEGFP molecules are switched off at 491 nm using a doughnut-shaped focal intensity (dashed blue line) so that the on-state is confined to a subdiffraction-sized region around the doughnut centre. Subsequent irradiation with 532-nm light makes the on-state molecules permanent by bleaching. Irradiation at 405 nm switches the off-state molecules back into the on-state, allowing the writing of another feature in subdiffraction proximity. Bottom, schematic of diffraction-limited writing. **b**, Conventional (left) and subdiffraction RESOLFT (middle) joined writing and reading in a layer of immobilized rsEGFP. The outlines of the corresponding 3×3 bit patterns were identical. The distance between two bleached spots was 250 nm in each case. Right, normalized line profiles of the fluorescence signal between the two arrows (solid red, RESOLFT; dashed blue, confocal). Scale bars, 1 μm .

TEMED. About 10 μ l of this solution was placed on a glass slide and a cover slip was pressed onto the sample to attain a thin layer. Custom MATLAB (The MathWorks) programs allowed automated generation of the voltages and signals for moving the sample and for generating the desired laser pulses. Images were also taken using the software Inspector (<http://www.inspector.de>).

RESOLFT set-up. We implemented a home-built confocal microscope with a normally focused beam for generating fluorescence plus a doughnut-shaped beam for switching rsEGFP off (both at 491 nm wavelength). The beams were circularly polarized, superimposed in the focal plane and applied sequentially. The 405 nm beam for switching rsEGFP on was also circularly polarized. The fluorescence emitted between 500–560 nm was imaged on the opening of a multi-mode fibre and detected by a counting avalanche photodiode. The same set-up was used for writing, which was most specific at 532 nm.

Full details of methods used are available in Supplementary Information.

Received 23 May; accepted 24 August 2011.

Published online 11 September 2011.

- Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
- Hell, S. W. & Kroug, M. Ground-state depletion fluorescence microscopy, a concept for breaking the diffraction resolution limit. *Appl. Phys. B* **60**, 495–497 (1995).
- Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
- Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl Acad. Sci. USA* **97**, 8206–8210 (2000).
- Hell, S. W. Toward fluorescence nanoscopy. *Nature Biotechnol.* **21**, 1347–1355 (2003).
- Hell, S. W. Microscopy and its focal switch. *Nature Methods* **6**, 24–32 (2009).
- Westphal, V. et al. Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science* **320**, 246–249 (2008).
- Hein, B., Willig, K. I. & Hell, S. W. Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell. *Proc. Natl Acad. Sci. USA* **105**, 14271–14276 (2008).
- Egeling, C. et al. Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature* **457**, 1159–1162 (2009).
- Nägerl, U. V., Willig, K. I., Hein, B., Hell, S. W. & Bonhoeffer, T. Live-cell imaging of dendritic spines by STED microscopy. *Proc. Natl Acad. Sci. USA* **105**, 18982–18987 (2008).
- Hell, S. W., Jakobs, S. & Kastrop, L. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Appl. Phys., A Mater. Sci. Process.* **77**, 859–860 (2003).
- Hell, S. W., Dyba, M. & Jakobs, S. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Opin. Neurobiol.* **14**, 599–609 (2004).
- Hell, S. W. Strategy for far-field optical imaging and writing without diffraction limit. *Phys. Lett.* **326**, 140–145 (2004).
- Gustafsson, M. G. L. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc. Natl Acad. Sci. USA* **102**, 13081–13086 (2005).
- Hell, S. W. in *Topics in Fluorescence Spectroscopy* Vol. 5 (ed. Lakowicz, J. R.) 361–422 (Plenum, 1997).
- Hofmann, M., Egeling, C., Jakobs, S. & Hell, S. W. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl Acad. Sci. USA* **102**, 17565–17569 (2005).
- Lukyanov, K. A. et al. Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J. Biol. Chem.* **275**, 25879–25882 (2000).
- Schwentker, M. A. et al. Wide-field subdiffraction RESOLFT microscopy using fluorescent protein photoswitching. *Microsc. Res. Tech.* **70**, 269–280 (2007).
- Ando, R., Mizuno, H. & Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **306**, 1370–1373 (2004).
- Dedecker, P. et al. Subdiffraction imaging through the selective donut-mode depletion of thermally stable photoswitchable fluorophores: numerical analysis and application to the fluorescent protein dronpa. *J. Am. Chem. Soc.* **129**, 16132–16141 (2007).
- Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* **143**, 1047–1058 (2010).
- Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
- Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–796 (2006).
- Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**, 355–358 (1997).
- Bossi, M., Foelling, J., Dyba, M., Westphal, V. & Hell, S. W. Breaking the diffraction resolution barrier in far-field microscopy by molecular optical bistability. *N. J. Phys.* **8**, 275 (2006).
- Scott, T. F., Kowalski, B. A., Sullivan, A. C., Bowman, C. N. & McLeod, R. R. Two-color single-photon photoinitiation and photoinhibition for subdiffraction photolithography. *Science* **324**, 913–917 (2009).
- Li, L., Gattass, R. R., Gershgoren, E., Hwang, H. & Fourkas, J. T. Achieving 1/20 resolution by one-color initiation and deactivation of polymerization. *Science* **324**, 910–913 (2009).
- Andrew, T. L., Tsai, H. Y. & Menon, R. Confining light to deep subwavelength dimensions to enable optical nanopatterning. *Science* **324**, 917–921 (2009).
- Fischer, J., Freymann, G. & Wegener, M. The materials challenge in diffraction-unlimited direct-laser-writing optical lithography. *Adv. Mater.* **22**, 3578–3582 (2010).
- Ormö, M. et al. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
- Andresen, M. et al. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc. Natl Acad. Sci. USA* **102**, 13070–13074 (2005).
- Andresen, M. et al. Structural basis for reversible photoswitching in Dronpa. *Proc. Natl Acad. Sci. USA* **104**, 13005–13009 (2007).
- Henderson, J. N., Ai, H. W., Campbell, R. E. & Remington, S. J. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proc. Natl Acad. Sci. USA* **104**, 6672–6677 (2007).
- Adam, V. et al. Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations. *Proc. Natl Acad. Sci. USA* **105**, 18343–18348 (2008).
- Brakemann, T. et al. Molecular basis of the light-driven switching of the photochromic fluorescent protein Padron. *J. Biol. Chem.* **285**, 14603–14609 (2010).
- Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. & Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* **73**, 2782–2790 (1997).
- Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
- Bizzarri, R. et al. Single amino acid replacement makes *Aequorea victoria* fluorescent proteins reversibly photoswitchable. *J. Am. Chem. Soc.* **132**, 85–95 (2010).
- Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Adam, V. et al. Data storage based on photochromic and photoconvertible fluorescent proteins. *J. Biotechnol.* **149**, 289–298 (2010).
- Vats, P. & Rothfield, L. Duplication and segregation of the actin (MreB) cytoskeleton during the prokaryotic cell cycle. *Proc. Natl Acad. Sci. USA* **104**, 17795–17800 (2007).
- Riedl, J. et al. Lifeact: a versatile marker to visualize F-actin. *Nature Methods* **5**, 605–607 (2008).
- Harris, K. M. & Kater, S. B. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annu. Rev. Neurosci.* **17**, 341–371 (1994).
- Hopt, A. & Neher, E. Highly nonlinear photodamage in two-photon fluorescence microscopy. *Biophys. J.* **80**, 2029–2036 (2001).
- Heintzmann, R., Jovin, T. M. & Cremer, C. Saturated patterned excitation microscopy—a concept for optical resolution improvement. *JOSA A* **19**, 1599–1609 (2002).
- Hell, S. W., Schmidt, R. & Egner, A. Diffraction-unlimited three-dimensional optical nanoscopy with opposing lenses. *Nature Photon.* **3**, 381–387 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Jethwa for careful reading and M. Andresen, T. Brakemann, S. Löbermann, R. Schmitz-Salue and A. C. Stiel for discussions and support, as well as T. Gilat and F. Voss (MPI of Neurobiology, Munich) for help with the slice culture preparation and A. Schönlé for adapting the software Inspector. We thank The Project Gutenberg for making Grimm's Fairy Tales available in electronic format, L. Rothfield (University of Connecticut Health Center) for providing the plasmid pLE7, R. Wedlich-Söldner (MPI of Biochemistry, Munich) for the lifeact-YFP construct and V. Stein (MPI of Neurobiology, Munich) for the virus protocol. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the DFG-Research Center for Molecular Physiology of the Brain (to S.J.) and by a Gottfried-Wilhelm-Leibniz prize of the DFG (to S.W.H.).

Author Contributions T.G., I.T., M.L., H.B., F.L.-C. performed research, I.T., M.L., T.G., H.B., C.E. set up the microscopes, N.T.U., K.I.W. prepared samples, M.L., T.G., I.T., K.I.W., S.J., S.W.H. analysed data, S.J., C.E., S.W.H. designed research. S.J., M.L., S.W.H. wrote the paper. All authors discussed the data and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.J. (sjakobs@gwdg.de) or S.W.H. (shell@gwdg.de).

DNA stretching by bacterial initiators promotes replication origin opening

Karl E. Duderstadt¹, Kevin Chuang² & James M. Berger^{1,2}

Many replication initiators form higher-order oligomers that process host replication origins to promote replisome formation. In addition to dedicated duplex-DNA-binding domains, cellular initiators possess AAA+ (ATPases associated with various cellular activities) elements that drive functions ranging from protein assembly to origin recognition. In bacteria, the AAA+ domain of the initiator DnaA has been proposed to assist in single-stranded DNA formation during origin melting. Here we show crystallographically and in solution that the ATP-dependent assembly of *Aquifex aeolicus* DnaA into a spiral oligomer creates a continuous surface that allows successive AAA+ domains to bind and extend single-stranded DNA segments. The mechanism of binding is unexpectedly similar to that of RecA, a homologous recombination factor, but it differs in that DnaA promotes a nucleic acid conformation that prevents pairing of a complementary strand. These findings, combined with strand-displacement assays, indicate that DnaA opens replication origins by a direct ATP-dependent stretching mechanism. Comparative studies reveal notable commonalities between the approach used by DnaA to engage DNA substrates and other, nucleic-acid-dependent, AAA+ systems.

All organisms depend on ring- and spiral-shaped ATPase assemblies to carry out essential processes ranging from proteolysis and membrane trafficking, to signalling events and nucleic acid transactions. DNA replication onset in cells reflects one such process, using ATP-dependent initiation factors to coordinate replisome assembly^{1,2}. Replication initiators of eukaryotes and prokaryotes contain AAA+-family ATPase domains, the activity of which is augmented by duplex-DNA-binding domains and specialized protein-protein interaction elements that assist with origin recognition and recruit specific replication factors^{3,4}. Although all AAA+ enzymes share a common structural core with RecA-type ATPases, together forming the additional strand catalytic glutamate (ASCE) supergroup of P-loop NTPases⁵, the molecular logic that allows a common nucleotidyl-hydrolase module to control the disparate activities of replication initiators, and ASCE proteins in general, is not understood.

In bacteria, replication initiation relies on the DnaA protein^{6–8}. In *Escherichia coli*, multiple DnaA molecules bind to the replication origin, *oriC*, through several duplex DNA-binding sites, forming a large nucleoprotein complex in the presence of ATP^{9–11}. With the aid of appropriate architectural proteins (such as integration host factor) and negatively supercoiled DNA, this complex subsequently melts an (A+T)-rich, DNA-unwinding element (DUE) located adjacent to the duplex DnaA binding sites^{12,13}. ATP also activates a secondary DNA-binding site within DnaA, thought to reside within the AAA+ domain, which engages single-stranded regions of the DUE to form a stable open complex^{12,14–16}. DnaA then collaborates with the bacterial helicase loader (DnaC in *E. coli*) to recruit two hexamers of the DnaB helicase to the origin and promote replisome assembly^{17–19}.

Although most AAA+ enzymes form closed-ring assemblies^{20,21}, structural studies have indicated that initiators and polymerase clamp-loaders form open-ring structures^{14,22–24}. Among initiator/loader systems, DnaA is particularly unusual in that it has been seen to oligomerize into a right-handed, spiral filament¹⁴. Two models have been proposed to explain how this structure might aid origin melting (Supplementary Fig. 1). In one, the wrapping of duplex DNA about a

DnaA superhelix would constrain a positive supercoil, generating compensatory negative writhe that could aid opening of the neighbouring DUE. In the other, the wrapped DnaA–DNA complex would serve as a nucleation centre, allowing DnaA protomers to engage directly and melt the DUE, possibly through the initiator's ATPase elements. Thus far, experimental evidence has supported both models^{9,14–16,25}, leaving open the question as to how DnaA catalyses origin melting. The relationship of this mechanism to other initiation systems, or to AAA+/ASCE proteins overall, is also unclear.

A DnaA–ssDNA crystal structure

To examine these issues, we set out to determine the structure of DnaA bound to single-stranded DNA (ssDNA). Using a truncation of *Aquifex aeolicus* DnaA consisting of the AAA+ and duplex-DNA-binding domains (which, like its *E. coli* counterpart^{16,17}, is active for both ATP-stimulated assembly and ssDNA binding²⁵), we first grew DNA-free crystals in the presence of Mg²⁺ and the non-hydrolysable ATP mimic AMPPCP¹⁴. DNA substrates were then soaked into these crystals under low-salt conditions (Methods). Data collection and phasing by molecular replacement revealed four DnaA protomers per asymmetric unit, arranged in a spiral configuration that propagates into a continuous protein helix by the action of crystal-symmetry elements (Fig. 1a, b), along with bound ssDNA. Of the multiple substrates screened (Methods), dA₁₂ yielded the highest-quality density (Supplementary Fig. 2a, b), and served as the best target for model building and refinement. The final structure, containing a DnaA:AMPPCP:Mg²⁺:dA₁₂ stoichiometry of 4:4:4:1, was refined to an $R_{\text{work}}/R_{\text{free}}$ of 24.9/26.8% at 3.35 Å resolution (Supplementary Table 1).

DnaA–ssDNA interactions

The overall arrangement of DnaA subunits in the helical assembly is highly similar to a DNA-free form reported previously (0.7 Å root mean squared deviation between all C α positions)¹⁴. AMPPCP•Mg²⁺ binds at the interface between neighbouring subunits, with the γ -phosphate of AMPPCP coordinated by catalytic amino acids from pairs of adjoining

¹Biophysics Graduate Group, University of California, Berkeley, Berkeley, California 94720, USA. ²Department of Molecular and Cell Biology, California Institute for Quantitative Biology, University of California, Berkeley, Berkeley, California 94720, USA.

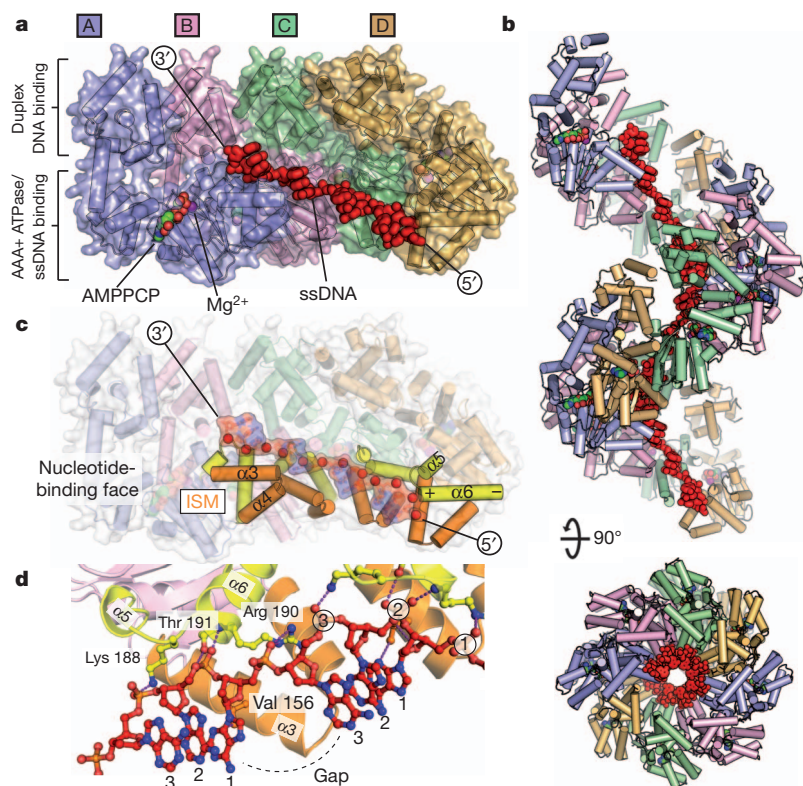


Figure 1 | The ATPase pore of assembled DnaA binds ssDNA. **a**, Side view of the asymmetric unit, with DnaA subunits differentially coloured. Single-stranded DNA is displayed as red sticks. AMPPCP and Mg^{2+} , bound to chain A, are shown as spheres coloured by element and in magenta, respectively; AMPPCP• Mg^{2+} bound to chains B–D is occluded in this view. **b**, Side and top views of oligomerized DnaA, reconstructed through crystal packing, showing 12 DnaA subunits and 3 strands of ssDNA. Colouring as in panel **a**. **c**, Side view of the DnaA tetramer with helices $\alpha 3/\alpha 4$ and $\alpha 5/\alpha 6$ highlighted in orange and yellow, respectively. ISM, initiator-specific motif. Single-stranded DNA is shown as a transparent stick-and-surface representation coloured by element; phosphates are further highlighted as red spheres. **d**, Protein–DNA contacts. Protein chains B (left) and C (right) are displayed with the same colouring as in panel **c**. Single-stranded DNA is coloured by element.

AAA+ domains. Single-stranded DNA associates exclusively with the AAA+ elements of the initiator, with each protomer binding three nucleotides of the dA_{12} strand (Fig. 1a). Almost all contacts are made through the phosphodiester backbone, exposing the DNA bases to solvent. Each trinucleotide segment adopts a B-form DNA conformation (Supplementary Fig. 3), with the bases between consecutive segments separated by large (~ 10 Å) gaps that extend the substrate by $\sim 50\%$ (Supplementary Table 2 and Supplementary Information).

DnaA binds ssDNA using just two pairs of helices, $\alpha 3/\alpha 4$ and $\alpha 5/\alpha 6$, both of which line the central channel of the protein assembly (Fig. 1c). The geometry of these two elements creates a single conduit along the length of the DnaA superhelix that allows substrate to traverse consecutive DnaA protomers. Interestingly, helices $\alpha 3/\alpha 4$ also comprise the initiator-specific motif, which both promotes filament formation^{14,19} and distinguishes DnaA as a member of the initiator clade of the AAA+ superfamily^{26,27}.

DnaA uses a simple network of interactions to coordinate ssDNA. The initiator-specific motif forms a shelf for each trinucleotide, in which a conserved hydrophobic residue, Val 156, forms van der Waals contacts with the sugar and base of the first nucleotide in the triplet (Fig. 1d). The central phosphate of each trinucleotide is bound by the electropositive, amino-terminal helix dipole of $\alpha 6$ and hydrogen bonded by Thr 191 (Fig. 1c, d). These contacts are flanked by two positively charged residues, Arg 190 and Lys 188, which make salt-bridge interactions with the phosphates of nucleotides 1 and 3, respectively. Notably, mutant initiators containing substitutions in these observed DNA-binding residues show reduced affinity for ssDNA in solution (Supplementary Fig. 4), confirming that the crystals captured a physiologically meaningful initiator state. Moreover, mutations of the same positions in *E. coli* DnaA (amino acids Arg 245, Lys 243 and Val 211) also disrupt ssDNA binding and origin melting¹⁵. Thus, the ssDNA engagement strategy seen here seems to be conserved across bacterial species.

Structural similarities between DnaA and RecA

In considering the assembly patterns of oligomeric ATPases, we were struck by the similarity of DnaA to one system in particular: the

homologous recombination protein, RecA. Although the cellular functions of these two proteins are fundamentally different (catalysis of DNA strand-exchange reactions versus replication origin melting and coordination of replisome assembly), both RecA and DnaA are predicated upon an ASCE ATPase fold^{27,28}. Like DnaA, RecA (and its Rad51/RadA orthologues) forms a helical assembly that engages DNA with its pore regions^{28–32}. These shared physical properties led us to undertake a more detailed comparison of RecA and DnaA. Of the multiple models available, the structure of a RecA oligomer bound to ssDNA³³, representing the presynaptic complex formed during the initial stages of homologous recombination, is globally most similar to the DnaA state we observe (Fig. 2a, b). As with DnaA, RecA contacts DNA almost exclusively through the phosphodiester backbone, which sits in the interior of a positively charged filament pore. Each RecA protomer binds three nucleotides in a B-DNA conformation, with the base stacking between each triplet interrupted such that ssDNA is extended ~ 1.5 -fold compared to a B-form duplex (Fig. 2c).

RecA and DnaA also exhibit some interesting and significant differences. A visual examination of each triplet shows that RecA uses a more extensive network of contacts for engaging ssDNA than does DnaA (Fig. 2d, e), burying twice as much surface area per triplet (318 Å² and 639 Å² for DnaA and RecA, respectively). This difference derives largely from an additional β -hairpin in RecA that fills the gap between each triplet and reinforces each three-base stack³³. Moreover, whereas two of the three nucleotides within each RecA triplet (positions 1 and 2) align well with those seen in DnaA, position 3 of the DnaA trinucleotide rotates away from the pore axis by $\sim 50^\circ$ (Fig. 2f). This difference skews consecutive DnaA triplets away from one another, disrupting the formation of a smoothly spiral arrangement as seen in RecA (Fig. 2c).

DNA extension is ATP- and assembly-dependent

The ability of RecA to stretch DNA to the extent observed crystallographically has been amply substantiated by various methodologies^{34–36}. Using these efforts as a guide, we set out to determine whether the DNA conformation that we observe bound to DnaA accurately represents the state of the substrate in solution. To accomplish this, we turned to a

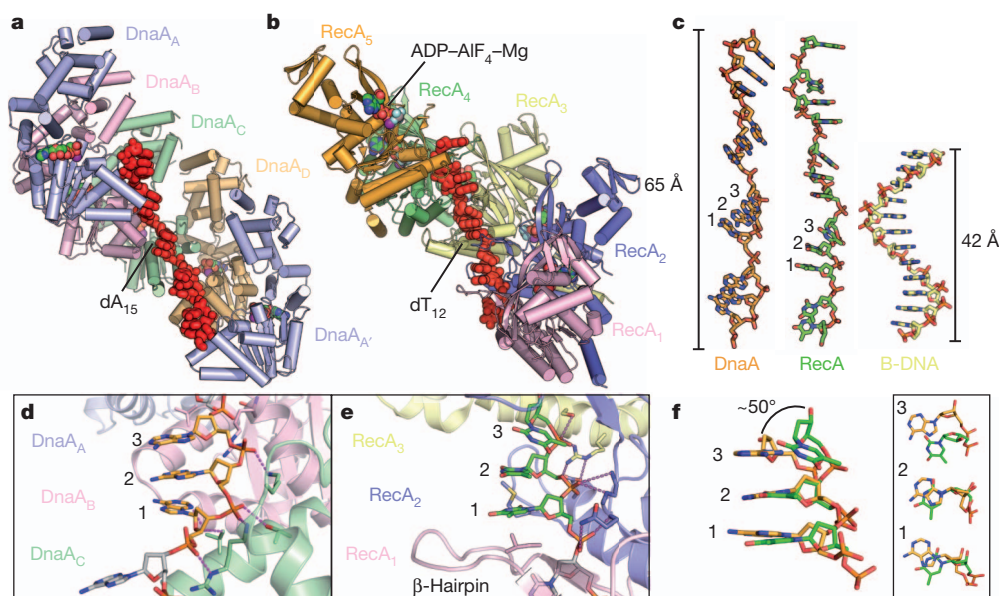


Figure 2 | DnaA engages ssDNA in a manner similar to RecA. **a**, View of a DnaA-AMPPCP-ssDNA pentamer (consisting of one full tetramer, as well as chain A (DnaA_{A'}) and its associated triplet from the adjacent asymmetric unit). AMPPCP•Mg²⁺ is shown as spheres coloured by atom; ssDNA as red sticks. **b**, View of a RecA-ADP-AlF₄-ssDNA pentamer (Protein Data Bank (PDB) accession 3CMW)³³. ADP•AlF₄•Mg²⁺ is shown as spheres coloured by atom;

ssDNA as red sticks. **c**, Comparison of ssDNA bound to DnaA (orange), RecA (green) and a strand of B-DNA (yellow). **d**, Close-up view of triplet bound to DnaA (chain C) with magenta dashed lines indicating key contacts. **e**, Close-up view of triplet bound to RecA (protomer 2) with magenta dashed lines indicating key contacts. **f**, Side (left) and top (right) views of the triplets displayed in **d** and **e** aligned with each other.

bulk-phase fluorescence resonance energy transfer (FRET)-based ssDNA extension assay analogous to single-molecule approaches applied to RecA³⁷. Using a poly-thymine DNA labelled with Cy3 and Cy5 (FR-dT₂₁) (Supplementary Table 3), we monitored changes in the length of ssDNA resulting from DnaA binding (Fig. 3a). Analogous studies were performed with RecA as a control. As both RecA and DnaA require ATP for formation of the oligomers observed in the structural models, we expected ATP-dependent extension to lead to

a loss of FRET signal. We tested for extension both in the presence of the ATP analogues ATPγS and ADP•BeF₃, to avoid complications that might arise from nucleotide hydrolysis, and in the presence of ADP, which is known to promote DnaA disassembly. Pronounced extension was observed only in the presence of the ATP analogues (Fig. 3b, c), and not with ADP. The lengths of ssDNA in the ATP-assembled states of both proteins, as calculated from the FRET data, were in close agreement with those observed in the crystal structures (Supplementary Table 6). Likewise, mutations in ssDNA-binding amino acids and residues required for DnaA assembly all significantly reduced ssDNA extension (Supplementary Fig. 6), demonstrating that this activity depends on substrate binding to the pore of an initiator oligomer that forms only when activated by ATP.

DnaA directly catalyses duplex melting

How replication origins are opened for replisome assembly is an important, unanswered question. Given the similarities between the ssDNA binding and extension activities of DnaA and RecA, we reasoned that the initiator might directly destabilize and disrupt DNA duplexes. This activity is a known property of RecA³⁸, albeit one that permits the recombinase to exchange DNA strands between target substrates actively^{30,33}.

To test this idea, we developed a DNA strand-displacement assay for DnaA. First, the initiator was incubated with a short duplex containing one fluorescently labelled strand. Unlabelled competitor strand was then added to capture any unwound species (Fig. 4a). Both ADP and ADP•BeF₃ were tested to determine whether initiator assembly affected the outcome of the experiment, as were DNAs of different lengths and stabilities. Analysis of the resultant products by gel electrophoresis shows that DnaA readily unwinds a 15mer duplex DNA of moderate stability ($T_m = 43^\circ\text{C}$) in the presence of the ATP mimic (Fig. 4b). By contrast, increasing the stability of the DNA substrate by ~30% (using a 20mer, $T_m = 55^\circ\text{C}$) weakens the unwinding activity of DnaA (Fig. 4b), while increasing DNA stability even further (30mer, $T_m = 62^\circ\text{C}$) abrogates melting completely (Supplementary Fig. 7a). Importantly, ADP did not support strand displacement, nor did ssDNA binding and DnaA assembly mutants (Supplementary Fig. 7b, c). These controls indicate that double-stranded-DNA melting is dependent not

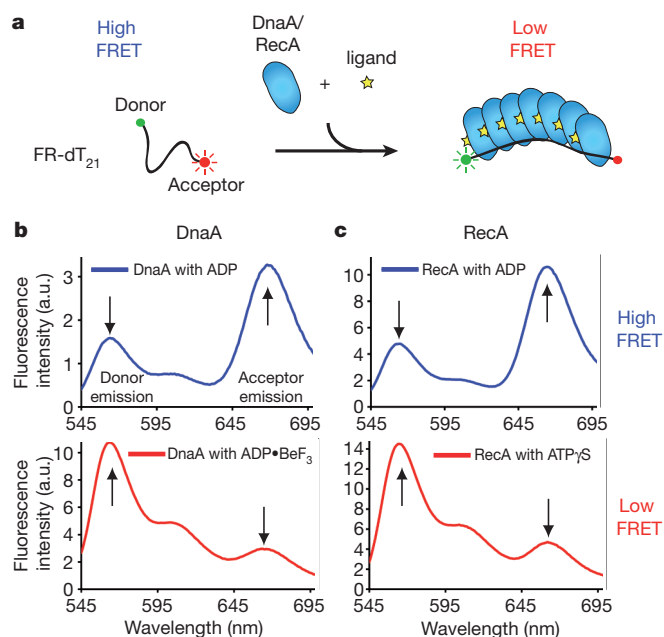


Figure 3 | DnaA extends ssDNA in solution. **a**, Cartoon of ssDNA extension assay. **b**, Emission scan (donor excitation) of FR-dT₂₁ in the presence of 10 μM DnaA with either ADP•BeF₃ (top) or ADP (bottom). **c**, Emission scan (donor excitation) of FR-dT₂₁ in the presence of 10 μM RecA with either ATPγS (top) or ADP (bottom). Reported transfer efficiencies and distances were calculated using donor emission as described in Methods. a.u., arbitrary units.

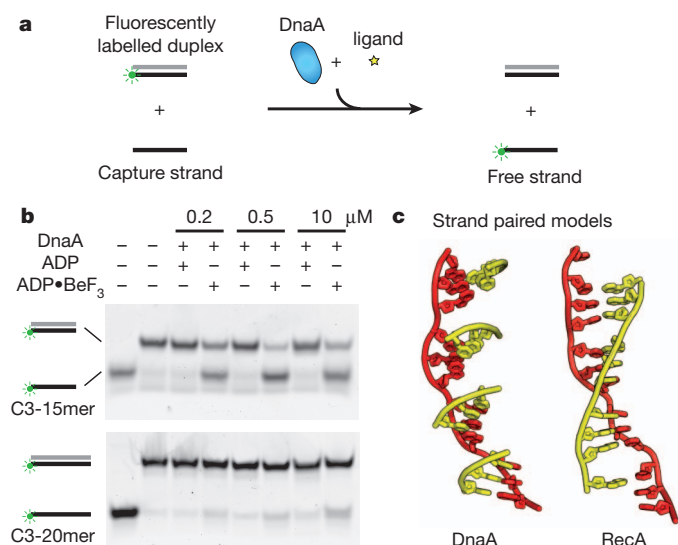


Figure 4 | DnaA directly melts duplex DNA. **a**, Schematic of strand displacement assay. The green circle represents the Cy3 fluorescent end-label used to follow the status of one DNA strand. Complementary strands of duplex substrates are coloured grey and black. **b**, Strand displacement assay conducted with 15mer and 20mer duplex substrates (C3-15mer and C3-20mer) in the presence and absence of different nucleotides. DnaA concentrations used are indicated above each lane. **c**, Left: cartoon model showing how complementary base triplets (yellow) would pair (in a B-DNA manner) with ssDNA bound to DnaA (red). The orientation of successive DnaA-bound triplets is such that it prevents the formation of a continuous base-paired strand favouring duplex separation. Right: same DNA view, but as seen in RecA, where triplets are oriented to allow pairing of an extended complementary strand to promote duplex formation and strand exchange (PDB accession 3CMX)³³.

only on formation of an assembled DnaA oligomer, but that the initiator is fine-tuned to specifically disrupt DNAs of modest stability.

One significant functional difference between RecA and DnaA is that the recombination protein can drive a true strand-exchange reaction; that is, in addition to displacing one strand of a duplex, RecA can also pair homologous ssDNA segments into a double-stranded molecule. By contrast, the function of DnaA is to separate double-stranded origin regions. Inspection of the RecA and DnaA complexes reveals a physical basis for these differing properties: in DnaA, successive trinucleotide elements are arranged in a state incompatible with the formation of a continuous duplex, whereas

ssDNA bound to RecA adopts a smoothly spiralled arrangement permitting the contiguous pairing of a complementary strand (Fig. 4c). This distinction arises primarily from the 50° rotation between the nucleotides at the third position of each triplet seen in the RecA and DnaA models (Fig. 2f). In DnaA, the orientation of this nucleotide appears to be stabilized by base stacking, whereas in RecA the β -hairpin insertion helps to sculpt the configuration of the DNA to create a contiguous base-pairing surface.

Implications for origin melting

Together, our findings present the strongest evidence yet that DnaA melts replication origins by directly assisting with the separation and sequestration of duplex DNA strands (Supplementary Fig. 1c). Notably, this activity does not contradict the demonstrated need for other factors capable of reshaping and/or destabilizing DNA (for example, integration host factor and negative supercoiling) during initiation^{12,13}. Rather, these elements probably help to promote DnaA assembly and prime the origin for melting by what otherwise would be an inefficient unwindase. In this view, the AAA+ domains of DnaA may first engage only one of the two strands of duplex DNA with their ssDNA binding elements (possibly at reported ssDNA or ATP-DnaA binding sites^{15,16}). In the presence of ATP, which triggers initiator assembly, subunit-subunit interactions would help to restructure the DNA backbone, stretching the contacted strand to facilitate melting. Re-annealing would be disfavoured by the non-contiguous arrangement of base triplets in the extended state (Fig. 4c). Future studies will be needed to define the specific order and effect of these events further.

We envision that the propensity of DnaA to open DNA could be adjusted in other bacterial species by strengthening or weakening the association of its ATPase domains with DNA and/or each other. An attractive feature of such a mechanism is that it is amenable to additional layers of control by changes to DUE sequence, superhelical density and co-resident architectural factors to ensure that a replication origin fires only when DnaA is both present and assembled properly. Such flexibility may have had a role in allowing DnaA to persist as the primary initiator in bacteria that have adapted to markedly different environmental niches.

Thematic patterns of substrate recognition in AAA+ ATPases

The mechanism by which DnaA coordinates ssDNA also comports well with findings in other replication initiation systems and with ASCE ATPases in general. For instance, many oligomeric RecA and

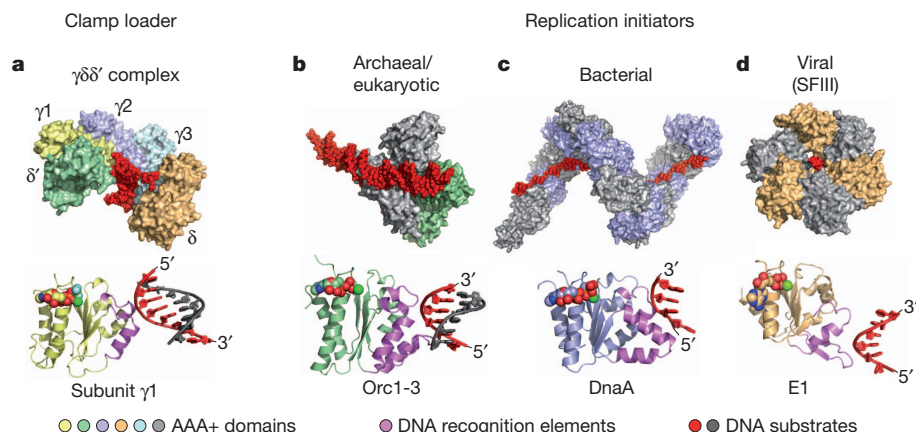


Figure 5 | Common DNA recognition strategies of AAA+ proteins. Structures of DNA-bound assemblies (top) and individual domains (bottom) for AAA+ proteins involved in replication. All recognize DNA using the same face of the AAA+ fold (violet; bottom). **a**, Bacterial clamp-loader ($\gamma\delta\delta'$) complex (AAA+ domains, differentially coloured) bound to primer-template DNA (PDB 3GLF)⁴¹. **b**, Archaeal initiators Orc1-1 (grey) and Orc1-3 (AAA+ domain, green) bound to origin DNA (PDB 2QBY)²⁶. **c**, Bacterial initiator DnaA (AAA+ domains, grey/blue) bound to ssDNA. **d**, Viral initiator/helicase E1 (AAA+ domains, orange/grey) bound to ssDNA (PDB 2GXA)⁴⁰. For all panels, DNA is shown as either red spheres (top) or as a red/grey cartoon (bottom). Nucleotide co-factors bound to AAA+ domains (bottom) are represented as spheres coloured by atom.

AAA+ enzymes bind substrate in the interior pore of a closed- or cracked-ring particle^{33,39–41}. DnaA follows this pattern. A comparison of DnaA to other, disparate nucleic-acid-dependent AAA+ systems—for example, polymerase clamp-loaders and processive helicases—further shows that these factors also associate with client substrates in a remarkably analogous manner, using the same face of the core $\alpha\beta\alpha$ ATP-binding fold to engage a short backbone stretch of their target DNAs (Fig. 5). For AAA+ proteins involved in initiation, these similar contact mechanisms have been differentially co-opted to assist with specific protein functions, ranging from the control of origin recognition (as seen in archaeal Orc1 proteins^{26,42}) to mediating processive DNA unwinding (viral superfamily 3 helicases^{40,43}). DnaA, with its ability to melt (but not translocate along) DNA, seems to use an intriguing mix of some of the activities exhibited by related initiation systems. Future efforts will be needed to determine how subtle differences in the position and nature of substrate-binding surfaces, combined with specific alterations in the assembly patterns of central AAA+ domains, endow such molecular motors and switches with their distinct biochemical properties.

METHODS SUMMARY

Detailed information regarding experimental methods, substrate sequences, binding constants, and FRET efficiencies and distances can be found in the Methods and in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 March; accepted 12 August 2011.

Published online 2 October 2011.

- Kornberg, A. & Baker, T. A. in *DNA Replication* (W. H. Freeman and Company, 1992).
- Stillman, B. Origin recognition and the chromosome cycle. *FEBS Lett.* **579**, 877–884 (2005).
- Duderstadt, K. E. & Berger, J. M. AAA+ ATPases in the initiation of DNA replication. *Crit. Rev. Biochem. Mol. Biol.* **43**, 163–187 (2008).
- Lee, D. G. & Bell, S. P. ATPase switches controlling DNA replication initiation. *Curr. Opin. Cell Biol.* **12**, 280–285 (2000).
- Leipe, D. D., Koonin, E. V. & Aravind, L. Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **333**, 781–815 (2003).
- Katayama, T., Ozaki, S., Keyamura, K. & Fujimitsu, K. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nature Rev. Microbiol.* **8**, 163–170 (2010).
- Kaguni, J. M. DnaA: controlling the initiation of bacterial DNA replication and more. *Annu. Rev. Microbiol.* **60**, 351–371 (2006).
- Leonard, A. C. & Grimwade, J. E. Regulating DnaA complex assembly: it is time to fill the gaps. *Curr. Opin. Microbiol.* **13**, 766–772 (2010).
- Fuller, R. S., Funnell, B. E. & Kornberg, A. The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38**, 889–900 (1984).
- Funnell, B. E., Baker, T. A. & Kornberg, A. *In vitro* assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* **262**, 10327–10334 (1987).
- Crooke, E., Thresher, R., Hwang, D. S., Griffith, J. & Kornberg, A. Replicatively active complexes of DnaA protein and the *Escherichia coli* chromosomal origin observed in the electron microscope. *J. Mol. Biol.* **233**, 16–24 (1993).
- Bramhill, D. & Kornberg, A. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* **52**, 743–755 (1988).
- Kowalski, D. & Eddy, M. J. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J.* **8**, 4335–4344 (1989).
- Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nature Struct. Mol. Biol.* **13**, 676–683 (2006).
- Ozaki, S. *et al.* A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin. *J. Biol. Chem.* **283**, 8351–8362 (2008).
- Speck, C. & Messer, W. Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* **20**, 1469–1476 (2001).
- Sutton, M. D., Carr, K. M., Vicente, M. & Kaguni, J. M. *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J. Biol. Chem.* **273**, 34255–34262 (1998).
- Fang, L., Davey, M. J. & O'Donnell, M. Replisome assembly at *oriC*, the replication origin of *E. coli*, reveals an explanation for initiation sites outside an origin. *Mol. Cell* **4**, 541–553 (1999).
- Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135**, 623–634 (2008).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells* **6**, 575–597 (2001).
- Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. Crystal structure of the processivity clamp loader gamma (γ) complex of *E. coli* DNA polymerase III. *Cell* **106**, 429–441 (2001).
- Speck, C., Chen, Z., Li, H. & Stillman, B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nature Struct. Mol. Biol.* **12**, 965–971 (2005).
- Clarey, M. G. *et al.* Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nature Struct. Mol. Biol.* **13**, 684–690 (2006).
- Duderstadt, K. E. *et al.* Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.* **285**, 28229–28239 (2010).
- Dueber, E. L., Corn, J. E., Bell, S. D. & Berger, J. M. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* **317**, 1210–1213 (2007).
- Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31 (2004).
- Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the *E. coli* recA protein monomer and polymer. *Nature* **355**, 318–325 (1992).
- Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
- Cox, M. M. Motoring along with the bacterial RecA protein. *Nature Rev. Mol. Cell Biol.* **8**, 127–138 (2007).
- Egelman, E. A common structural core in proteins active in DNA recombination and replication. *Trends Biochem. Sci.* **25**, 179–182 (2000).
- Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nature Struct. Mol. Biol.* **11**, 791–796 (2004).
- Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489–494 (2008).
- Stasiak, A. & Di Capua, E. The helicity of DNA in complexes with recA protein. *Nature* **299**, 185–186 (1982).
- Galletto, R., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature* **443**, 875–878 (2006).
- Nishinaka, T., Ito, Y., Yokoyama, S. & Shibata, T. An extended DNA structure through deoxyribose-base stacking induced by RecA protein. *Proc. Natl Acad. Sci. USA* **94**, 6623–6628 (1997).
- Joo, C. *et al.* Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515–527 (2006).
- Bianchi, M., Riboli, B. & Magni, G. *E. coli* recA protein possesses a strand separating activity on short duplex DNAs. *EMBO J.* **4**, 3025–3030 (1985).
- Thomsen, N. D. & Berger, J. M. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**, 523–534 (2009).
- Enemark, E. J. & Joshua-Tor, L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**, 270–275 (2006).
- Simonetta, K. R. *et al.* The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* **137**, 659–671 (2009).
- Gaudier, M., Schuwirth, B. S., Westcott, S. L. & Wigley, D. B. Structural basis of DNA replication origin recognition by an ORC protein. *Science* **317**, 1213–1216 (2007).
- Liu, X., Schuck, S. & Stenlund, A. Adjacent residues in the E1 initiator β -hairpin define different roles of the beta-hairpin in Ori melting, helicase loading, and helicase activity. *Mol. Cell* **25**, 825–837 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank K. Drlica, J. Keck, T. Murray and the Berger laboratory for helpful comments, and M. M. Cox for his contribution of RecA protein. This work was supported by the NIGMS (GM071747) and the National Institute of Health Molecular Biophysics Training Grant T32 GM008295.

Author Contributions K.E.D. and J.M.B. designed the experiments, analysed the data and wrote the paper. Protein purification, crystallization and ssDNA binding assays were performed by K.C. and K.E.D. K.E.D. performed the other experiments.

Author Information Coordinates have been deposited in the RSCB Protein Data Bank under the accession number 3R8F. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.B. (jmberger@berkeley.edu).

METHODS

Expression and purification of DnaA. Residues 76–399 of *A. aeolicus* DnaA (containing the AAA+ and duplex-DNA-binding regions) were expressed as a TEV-protease cleavable His₆-MBP fusion and purified as previously described²⁵. As a final purification step, untagged DnaA proteins (from TEV cleavage) were run over an S-200 size-exclusion column (GE) in gel-filtration buffer (50 mM HEPES pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 5 mM MgCl₂, 100 μM ADP). Monomeric species were pooled, concentrated and flash-frozen for storage at –80 °C. For mutagenesis studies, changes were introduced into the His₆-MBP–DnaA construct using QuickChange (Stratagene).

Crystallization and DNA soaking. After gel filtration of DnaA in crystallization buffer (20 mM HEPES pH 7.5, 250 mM KCl, 250 mM KBr, 10% (v/v) glycerol, 10 mM MgCl₂, 100 μM AMPPCP), monomeric species were pooled, concentrated to 10 mg ml^{–1} at 4 °C, and flash-frozen for storage at –80 °C. Crystallization by hanging-drop vapour diffusion was performed by mixing 1.3 μl of freshly thawed DnaA in crystallization buffer and 1 μl of well solution (15–35 mM sodium cacodylate pH 6.5, 26% 1,2-propanediol and 1–2% PEG 2000 MME) at 18 °C. Large rod-like crystals appeared within 1 to 2 weeks and reached maximal size around 3 weeks. Crystals were transferred by looping to a low-salt soaking solution (20 mM HEPES pH 7.5, 30 mM sodium cacodylate pH 6.5, 10% (v/v) glycerol, 10 mM MgCl₂, 26% 1,2-propanediol, 2.5% PEG 2000 MME and 200 μM AMPPCP) containing 5 mM ssDNA. After 6 h, crystals were looped and transferred to a second drop of soaking solution containing 5 mM ssDNA, and left overnight to ensure both complete removal of remaining salt and to allow time for binding. The crystals were then looped and flash frozen in liquid nitrogen in preparation for data collection. Previous biochemical studies revealed no apparent sequence preference for ssDNA by *A. aeolicus* DnaA²⁵, so DNAs of various sequences and lengths were all individually tested (dT_n (*n* = 3–12) and dA_n (*n* = 3–12), Elim Biopharmaceuticals). Data collection and structure determination revealed that dA₁₂ generated the strongest electron density, although similar, albeit weaker and less connected, density was observed for dT oligonucleotides and smaller dA substrates.

Data collection and structure determination. Data were collected at Beamline 8.3.1 at the Advanced Light Source (ALS)⁴⁴ and processed using HKL-2000⁴⁵. Crystals belong to the space group P2₁2₁2₁, with dA₁₂ soaked crystals having unit cell dimensions *a* = 99.8 Å, *b* = 114.2 Å and *c* = 201.3 Å (Supplementary Table 1). Data were phased using molecular replacement as implemented in PHENIX⁴⁶, using a DNA-free DnaA tetramer as a search model (PDB 2HCB)¹⁴. Initial *F*_o – *F*_c electron density maps containing clear density for DNA were generated using rigid-body and grouped B-factor refinement with PHENIX⁴⁶. Further refinement was conducted using multibody, non-crystallographic-symmetry (NCS)-restrained, simulated annealing in PHENIX⁴⁶, fourfold multibody NCS averaging with a custom solvent mask (including the region of DNA binding), density modification using resolve⁴⁷, and manual model building in COOT⁴⁸. During the final stages of refinement, fourfold multibody NCS and secondary structure restraints were retained for AMPPCP and the entire protein except residues 255–265, which differed between chains as a result of crystal packing interactions. Composite, simulated-annealing omit maps generated with CNS⁴⁹ were used as a guide for building with COOT. DNA and waters were manually added to the model, and final rounds of refinement with PHENIX were conducted with grouped B-factor modelling, as well as NCS restraints and TLS modelling of individual protein domains (comprising three TLS groups in total: the AAA+-core (amino acids 76–241) plus AMPPCP; the AAA+ α-helical 'lid' (amino acids 242–254 and 266–241); and the duplex-DNA binding domain (amino acids 291–399)). All panels of figures with renderings of structures and electron density were prepared with PyMol⁵⁰.

The final model contains one DnaA tetramer bound to one dA₁₂ per asymmetric unit. A clear 5' or 3' break between successive dA₁₂ substrates was not present in the electron density, indicating that during the soaking procedure, different DNA molecules bound in multiple registers to consecutive DnaA protomers throughout the crystal. Accordingly, a terminal 5' phosphate, which was not present in the substrate used for soaking, was added to the modelled dA₁₂ DNA.

Polarity of DNA binding. During refinement, DNA was initially modelled independently into the DnaA pore with each of the two possible polarities. Compared to the 5' to 3' polarity presented in the paper, refinement of the model with the dA₁₂ substrate running 3' to 5' (from the arginine finger side to the nucleotide-binding face of a DnaA protomer) resulted in only marginally higher *R*_{work} and *R*_{free} values (~0.1%), but also the appearance of off-model positive difference density and on-model negative difference density in *F*_o – *F*_c maps (the model as presented in the paper did not display such features). Simultaneous refinement with two DNAs, each at half occupancy, with opposing polarities of the dA₁₂

substrate resulted in ~0.3% higher *R*_{work} and *R*_{free} values, and again showed unfavourable difference density in *F*_o – *F*_c maps.

Recognizing that these differences, although consistent with our build, were subtle and did not definitively resolve the ssDNA binding polarity to DnaA, we set out to test our assignment further. To this end, we designed and had synthesized (by Trilink BioTechnologies) two specialized, di-adenosyl nucleotide substrates that would give rise to a clear distinction in binding orientation: 5'-p(Br-A)pAp and 5'-p(ε-A)pAp, where 'Br-A' indicates a bromo-deoxyadenosine label, 'ε-A' indicates an etheno-deoxyadenosine label, and 'p' indicates a phosphate moiety. Soaking of crystals with these dinucleotide substrates was performed as described for ssDNA substrates (note that in our soaking trials with oligonucleotides as short as dA₃, we observed density associated with DnaA protomers consistent with that seen for the trinucleotide repeats when using dA₁₂). Unfortunately, data collected with the ε-A-substituted oligonucleotide yielded maps with density for dinucleotides bound to each monomer but additional density for the EthenoA was not clearly visible, probably due to the low ~3.4 Å resolution limit of the DnaA crystals. At the same time, SAD data sets collected with the Br-A-substituted oligonucleotide did not yield useful maps, due to the weak diffraction of the crystals (and accompanying radiation damage as we attempted to maximize data signal-to-noise at the bromine absorption maximum⁵¹), to incomplete bromine labelling, or both. We note that we carried out soaks with longer Br-dA-labelled (and Br-dU-labelled) oligonucleotides, but these efforts were not successful, again because of weak diffraction. Additional experiments to test the orientation (for example, using labelled oligo/protein pairs and FRET) were considered, but ruled out due to the small binding site size for substrate DNA, and an inability to find a suitable pair of labelling sites that could report on differing binding orientations.

As a consequence, although our data are supportive of the polarity presented in our model, we cannot definitively rule out the possibility that ssDNA might also be binding to DnaA in the crystal in an opposing direction. Nonetheless, several findings support the idea that DnaA binds ssDNA in a defined orientation that is consistent with the direction suggested here. For example, following nucleoprotein complex formation on *oriC*, DnaA melts (A+T)-rich regions in the DUE¹²; two independent reports have found that *E. coli* DnaA binds specifically to only one strand (the so-called 'top' strand) of the DUE during this process^{15,16}. The importance of DnaA binding polarity becomes clear during the next stage of initiation, when the DnaB helicase is loaded. Modelling studies based on the known DnaB translocation polarity (5' to 3') and known pairwise interactions between DnaB, DnaC and DnaA, have suggested that top-strand loading involves a direct interaction between DnaA and DnaC that has been observed biochemically and depends on the AAA+ domains of the two proteins¹⁹. Because AAA+ domains assemble with a defined orientation, in which the arginine finger face of one protomer points into the nucleotide binding face of a second subunit, it follows that DnaA molecules probably position themselves on the top strand with only one of their two AAA+ domain surfaces presented to DnaC. Although the polarity of the DnaA–DnaC interaction has not been established, a mutation on the arginine finger face of the *E. coli* DnaA AAA+ domain, R281A, is reported to disrupt helicase loading, but not *oriC* melting⁵²; this finding indicates that DnaA interacts with DnaC using its arginine-finger face. In our structures, the 5' end of the modelled DNA resides near the arginine finger face of DnaA, a configuration consistent with these data.

ssDNA extension assay. Extension of dT₂₁ oligonucleotides labelled with Cy3 and Cy5 (FR-dT₂₁) by DnaA was monitored by FRET using a FluoroMax-4 (Horiba Jobin Yvon) spectrofluorimeter. Measurements were carried out at 25 °C in 20 μl with 25 nM of FR-dT₂₁ and either 10 μM of DnaA in DnaA extension buffer (50 mM HEPES pH 7.5, 125 mM KCl, 2% (v/v) glycerol, 10 mM MgCl₂ and 2 mM ADP or ADP•BeF₃) or 10 μM of RecA in RecA extension buffer (25 mM Tris-acetate pH 7.5, 100 mM Na-acetate, 10 mM Mg-acetate, 1 mM DTT and 2 mM ADP or ATPγS). Emission scans from 545 to 700 nm were collected with excitation of Cy3 at 530 nm, divided by the excitation intensity, and then corrected for the wavelength-dependent sensitivity of the detector. FRET efficiencies and distances were determined by comparing the Cy3 fluorescence from the doubly labelled substrate (FR-dT₂₁) with the Cy3 fluorescence from a substrate only having a Cy3 label (C3-dT₂₁) under the same conditions.

Influence of proteins on dye behaviour. To ensure that all influences on dye behaviour were properly considered when processing the FRET data from the DNA extension assay (Fig. 3), the fluorescence and absorbance of each dye was monitored independently for each experimental condition. Emission and absorbance scans of substrates labelled only with the Cy3 donor (C3-dT₂₁) were collected in buffer alone, and with protein in the presence of ATP mimics or ADP (Supplementary Fig. 8c (panels ii and iii) and d (panels ii and iii)). Emission scans revealed pronounced protein- and nucleotide-dependent enhancement of donor fluorescence, but negligible differences in donor absorbance. Similar, but less significant, effects have been observed previously for RecA at a concentration

of 1 μM (comparable to our working concentration of 10 μM)³⁷. A similar enhancement in acceptor fluorescence (and lack of effect on acceptor absorbance) also was observed in the doubly labelled substrate, FR-dT₂₁ (Supplementary Fig. 8c (panels iv and v) and d (panels iv and v)). These effects are not surprising, as the spectral properties of fluorescent dyes are known to undergo marked variation depending on chemical environment^{53,54}. However, these controls also indicated that we needed to take into account additional corrections to obtain accurate distance measurements. In particular, the changes in donor fluorescence, but not donor absorbance, were indicative of changes in the donor quantum yield (Φ_D), which is used to calculate R_0 (\AA), the distance corresponding to a FRET efficiency of 50%: $R_0 = 8.79 \times 10^{-5} (J\kappa^2 n^{-4} \Phi_D)^{1/6}$ where J is the spectral overlap between the donor emission and acceptor absorption; κ^2 is a geometric factor that depends on the orientation of donor and acceptor; and n is the refractive index of the medium between donor and acceptor⁵⁴.

To determine the donor quantum yield under different experimental conditions, we used rhodamine 6G as a standard for calibration, with an assumed quantum yield of 0.95 in EtOH⁵⁵. To calculate the quantum yields seen in Supplementary Table 5, we collected the fluorescence and absorbance of the donor-only labelled substrate (C3-dT₂₁) in the RecA and DnaA buffers. We then used these data to determine ratios between the integrated fluorescence and absorbance, while correcting for the fractional absorbance at the excitation wavelengths used. Fluorescein in 0.1 M NaOH (known to have a quantum yield of 0.95 (ref. 56)) was also measured as a control. To ensure reliable readings, all absorbance measurements were conducted with 1 μM of the donor-only labelled substrate, either alone or in the presence of 10 μM of the indicated protein. All emission measurements were conducted with 25 nM of the donor-only labelled substrate, either alone or in the presence of 10 μM of the indicated protein. Because the presence of protein had no influence on dye absorbance (Supplementary Fig. 8c (panels iii and v)), the quantum yield of the donor in the presence of different proteins was determined simply by using its value in buffer, and multiplying by the observed changes in fluorescence. R_0 values were then calculated for each sample using the corresponding values for quantum yield.

Determination of FRET efficiencies and DNA length. To determine the efficiency of transfer (E) from the FRET data collected using the DNA extension assay (Fig. 3 and Supplementary Figs 6, 8 and 9), the emission of the donor from the donor-only labelled substrate (C3-dT₂₁, F_D) was compared to the emission of the donor from the doubly labelled substrate FR-dT₂₁ (F_{DA}) under equivalent experimental conditions as follows: $E = 1 - \frac{F_{DA}}{F_D}$ (ref. 57). The efficiencies for different samples can be found in Supplementary Table 5. Solution distances were subsequently obtained using the relation $R = R_0 \left[\frac{1-E}{E} \right]^{1/6}$ (ref. 54), the values of which can also be found in Supplementary Table 5.

DNA strand displacement assay. The DnaA-dependent displacement of single strands from duplex DNA was monitored using a Cy3 label on one of two strands (the 'bottom' strand, Supplementary Table 3). All measurements were carried out at 25 °C in 80 μl of binding buffer containing 50 mM HEPES pH 7.5, 125 mM KCl, 2% (v/v) glycerol, 10 mM MgCl₂, 0.1 mg ml⁻¹ bovine serum albumin, 1 mM DTT and 2 mM ADP or ADP•BeF₃ (a non-hydrolysable ATP analogue that mimics the properties of ATP^{25,58}). After a short 2-min incubation of 25 nM duplex DNA with various DnaA concentrations (Fig. 4), 50 nM of unlabelled bottom strand was added for an additional 30 min to capture displaced top strands. After quenching with 10 \times stop buffer containing 200 mM EDTA, 10 mg ml⁻¹ proteinase K and 4% (v/v) SDS, displaced strands were separated on native polyacrylamide gels in Tris/boric acid/EDTA (TBE) buffer and visualized using a Molecular Dynamics Typhoon. The time dependence of DNA strand displacement by DnaA can be found in Supplementary Fig. 11. Sequences of substrates used can be found in Supplementary Table 3.

ssDNA binding assay. Binding of 5' fluorescein-labelled dT₂₅ oligonucleotides (F-dT₂₅, Supplementary Table 3) to DnaA was monitored by fluorescence polarization

using a Victor 3V (Perkin Elmer) multi-label plate reader (Supplementary Fig. 4a). Measurements were carried out at 25 °C in 20 μl of binding buffer containing 50 mM HEPES pH 7.5, 125 mM KCl, 2% (v/v) glycerol, 10 mM MgCl₂, 0.1 mg ml⁻¹ bovine serum albumin, 1 mM DTT and 2 mM ADP•BeF₃. The concentration of F-dT₂₅ was held constant at 10 nM while the concentration of DnaA was varied. All data points represent the average of three independent measurements, with error bars representing the standard deviation between measurements. Binding curves were fit to the Hill equation to obtain $K_{d,app}$ values (Supplementary Table 4) as described previously²⁵.

Oligomerization characteristics of ssDNA-binding mutants. To confirm that ssDNA binding mutations did not affect the ATP-dependent oligomerization properties of DnaA, we used a previously established glutaraldehyde-crosslinking assay²⁵. Crosslinking was performed by incubating 50 μg ml⁻¹ of various *A. aeolicus* DnaA proteins in 80 μl of a reaction buffer (50 mM HEPES pH 7.5, 10% (v/v) glycerol, 125 mM KCl, 5 mM MgCl₂, 2 mM DTT) containing 2 mM ADP•BeF₃ at 25 °C for 5 min. Glutaraldehyde (Polysciences Inc.) was then added to 1 mM final concentration using 8.8 μl of a 10 mM stock. Reactions were incubated at 25 °C for an additional 1 min before quenching with 8 μl of 200 mM glycine followed by the addition of 30 μl of gel loading buffer (100 mM Tris pH 6.8, 24% (v/v) glycerol, 8% (w/v) SDS, 200 mM DTT, 0.02% (w/v) bromophenol blue). Crosslinked proteins were loaded in a volume of 15 μl and separated on denaturing 4.5% polyacrylamide gels (80:1 acrylamide:bisacrylamide) in 0.1 M sodium phosphate, 0.1% SDS buffer (pH 7.2)^{59,60}, and visualized by silver staining (Supplementary Fig. 4b).

44. MacDowell, A. A. *et al.* Suite of three protein crystallography beamlines with single superconducting bend magnet as the source. *J. Synchrotron Radiat.* **11**, 447–455 (2004).
45. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* 307–326 (Academic Press, 1997).
46. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
47. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
49. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
50. DeLano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).
51. Ennifar, E., Carpentier, P., Ferrer, J. L., Walter, P. & Dumas, P. X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr. D* **58**, 1262–1268 (2002).
52. Felczak, M. M. & Kaguni, J. M. The box VII motif of *Escherichia coli* DnaA protein is required for DnaA oligomerization at the *E. coli* replication origin. *J. Biol. Chem.* **279**, 51156–51162 (2004).
53. Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J. & Waggoner, A. S. Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug. Chem.* **4**, 105–111 (1993).
54. Cantor, C. R. & Schimmel, P. R. in *Biophysical Chemistry: Part II: Techniques for the study of biological structure and function* 846 (W. H. Freeman and Company, 1980).
55. Magde, D., Wong, R. & Seybold, P. G. Fluorescence quantum yields and their relation to lifetimes of rhodamine 6G and fluorescein in nine solvents: improved absolute standards for quantum yields. *Photochem. Photobiol.* **75**, 327–334 (2002).
56. Lakowicz, J. R. in *Principles of Fluorescence Spectroscopy* 2nd edn. (Kluwer/Plenum, 1999).
57. Clegg, R. M. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.* **211**, 353–388 (1992).
58. Petsko, G. A. Chemistry and biology. *Proc. Natl Acad. Sci. USA* **97**, 538–540 (2000).
59. Crisona, N. J. & Cozzarelli, N. R. Alteration of *Escherichia coli* topoisomerase IV conformation upon enzyme binding to positively supercoiled DNA. *J. Biol. Chem.* **281**, 18927–18932 (2006).
60. Weber, K. & Osborn, M. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Biol. Chem.* **244**, 4406–4412 (1969).

Low-Mach-number turbulence in interstellar gas revealed by radio polarization gradients

B. M. Gaensler¹, M. Haverkorn^{2,3,4}, B. Burkhart⁵, K. J. Newton-McGee^{1,6}, R. D. Ekers⁶, A. Lazarian⁵, N. M. McClure-Griffiths⁶, T. Robishaw¹, J. M. Dickey⁷ & A. J. Green¹

The interstellar medium of the Milky Way is multiphase¹, magnetized² and turbulent³. Turbulence in the interstellar medium produces a global cascade of random gas motions, spanning scales ranging from 100 parsecs to 1,000 kilometres (ref. 4). Fundamental parameters of interstellar turbulence such as the sonic Mach number (the speed of sound) have been difficult to determine, because observations have lacked the sensitivity and resolution to image the small-scale structure associated with turbulent motion^{5–7}. Observations of linear polarization and Faraday rotation in radio emission from the Milky Way have identified unusual polarized structures that often have no counterparts in the total radiation intensity or at other wavelengths^{8–12}, and whose physical significance has been unclear^{13–15}. Here we report that the gradient of the Stokes vector (Q , U), where Q and U are parameters describing the polarization state of radiation, provides an image of magnetized turbulence in diffuse, ionized gas, manifested as a complex filamentary web of discontinuities in gas density and magnetic field. Through comparison with simulations, we demonstrate that turbulence in the warm, ionized medium has a relatively low sonic

Mach number, $M_s \lesssim 2$. The development of statistical tools for the analysis of polarization gradients will allow accurate determinations of the Mach number, Reynolds number and magnetic field strength in interstellar turbulence over a wide range of conditions.

We consider radio-continuum images of an 18-deg² patch^{11,16} of the Galactic plane, observed with the Australia Telescope Compact Array (ATCA) at a frequency of 1.4 GHz. Data were simultaneously recorded in total intensity (Stokes parameter I) and in linear polarization (Stokes parameters Q and U). The Stokes I image (Fig. 1) shows a typical distribution of radio emission, consisting of supernova-remnant shells, ionized regions around massive stars (H II regions) and unresolved distant radio sources. However, the corresponding images of Q , U and the linearly polarized intensity $P \equiv (Q^2 + U^2)^{1/2}$ in Fig. 1 are filled with complex structure that bears little resemblance to the Stokes I image, as has also been seen in many other polarimetric observations at radio frequencies^{8,9,12}. The intensity variations seen in Q , U and P are the result of small-scale angular structure in the Faraday rotation induced by ionized gas⁸, and are thus an indirect representation of

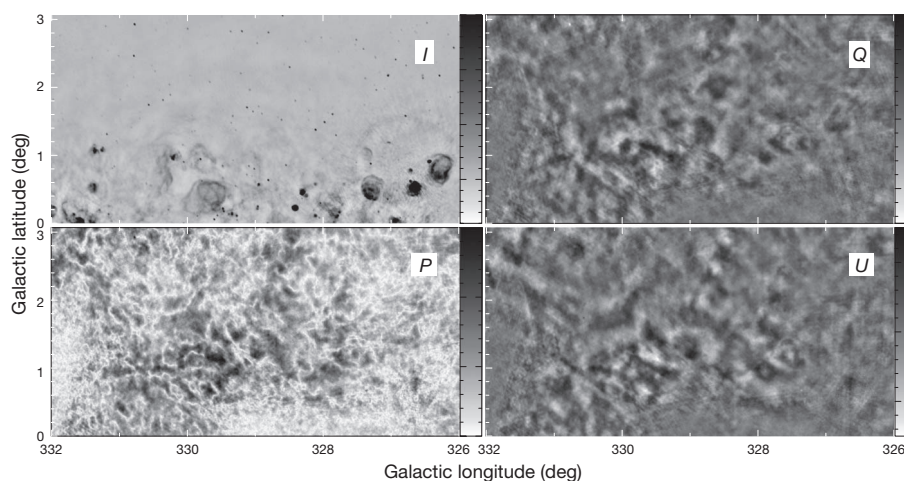


Figure 1 | Total intensity (I) and linearly polarized intensity (Q , U , P) for an 18-deg² region of the Southern Galactic Plane Survey²⁹. All four images were generated¹¹ from a set of observations¹⁶ taken at the ATCA over the period 1997 April to 1998 April using a 96-MHz bandwidth centred on an observing frequency of 1,384 MHz. The field is a mosaic of 190 pointings each with a total integration time of 20 min, resulting in an approximately uniform sensitivity, over most of the field, of 0.8 mJy per beam (Stokes I) or 0.55 mJy per beam (Stokes Q and U) at an angular resolution of 75 arcsec ($1 \text{ Jy} \equiv 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$). The scale for each image is shown on the right of each panel. The Stokes I image is displayed over a range of -40 to $+150$ mJy per beam (each interval corresponds to 10 mJy per beam). Because the ATCA is an interferometer, it is

not sensitive to structure on angular scales larger than 35 arcmin. Faint wisps can be seen, corresponding to the sharp edges of large-scale structures. However, the bulk of the smooth radio emission from Galactic cosmic rays is not detected. Imaging artefacts in the form of grating rings and radial streaks can be seen around a few very bright sources, but these regions were not used in our statistical analysis. The Stokes Q and U images are displayed over a range of -15 to $+15$ mJy per beam (interval, 2 mJy per beam), and the P image covers a range of 0 to 15 mJy per beam (interval, 1 mJy per beam). Almost none of the structure seen in Q , U and P has any correspondence with any emission seen in Stokes I ; the mottled structure results from spatial fluctuations in Faraday rotation in the ISM.

¹Sydney Institute for Astronomy, School of Physics, The University of Sydney, New South Wales 2006, Australia. ²ASTRON, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands. ³Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands. ⁴Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands. ⁵Astronomy Department, University of Wisconsin, Madison, 475 North Charter Street, Madison, Wisconsin 53711, USA. ⁶Australia Telescope National Facility, CSIRO Astronomy and Space Science, PO Box 76, Epping, New South Wales 1710, Australia. ⁷School of Mathematics and Physics, University of Tasmania, Private Bag 37, Hobart, Tasmania 7001, Australia.

turbulent fluctuations in the free-electron density and magnetic field throughout the interstellar medium¹⁵ (ISM).

A limitation of previous studies is that they usually interpreted the data in terms of the amplitude, P , and/or the angle, $\theta \equiv (1/2)\tan^{-1}(U/Q)$, of the complex Stokes vector $\mathbf{P} \equiv (Q, U)$. However, neither polarization amplitude nor polarization angle is preserved under arbitrary translations and rotations in the Q – U plane. These can result from one or more of a smooth distribution of intervening polarized emission, a uniform screen of foreground Faraday rotation, and the effects of missing large-scale structure in an interferometric data set. In the most general case, we are thus forced to conclude that the observed values of P and θ do not have any physical significance, and that only measurements of quantities that are both translationally and rotationally invariant in the Q – U plane can provide insight into the physical conditions that produce the observed polarization distribution.

The simplest such quantity is the spatial gradient of \mathbf{P} , that is, the rate at which the polarization vector traces out a trajectory in the Q – U plane as a function of position on the sky. The magnitude of the gradient is unaffected by rotation and translation, and so has the potential to reveal properties of the polarization distribution that might otherwise be hidden by excess foreground emission or Faraday rotation, or in data sets from which large-scale structure is missing (as is the case for the data shown in Fig. 1). The magnitude of the polarization gradient is

$$|\nabla\mathbf{P}| = \sqrt{\left(\frac{\partial Q}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial Q}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2} \quad (1)$$

The expression in equation (1) can be calculated simply, and the corresponding image of $|\nabla\mathbf{P}|$ (Fig. 2) reveals a complex network of tangled filaments. In particular, all regions in which $|\nabla\mathbf{P}|$ is high consist of elongated, narrow structures rather than extended patches. In the inset of Fig. 2, we plot the direction of $\nabla\mathbf{P}$ for a small subregion of the image, demonstrating that $\nabla\mathbf{P}$ changes most rapidly along directions oriented perpendicular to the filaments. We can explore the

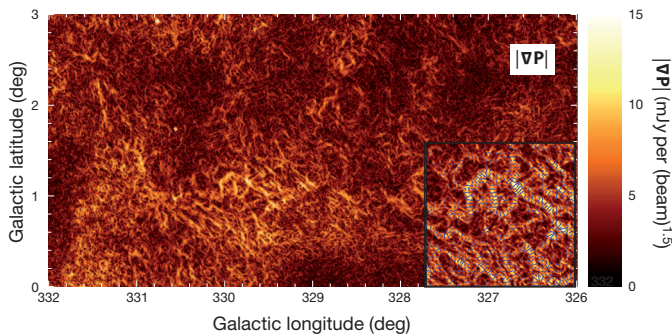


Figure 2 $|\nabla\mathbf{P}|$ for an 18-deg² region of the Southern Galactic Plane Survey.

$|\nabla\mathbf{P}|$ has been derived by applying equation (1) to the Q and U images from Fig. 1; note that $|\nabla\mathbf{P}|$ cannot be constructed from the scalar quantity $P \equiv (Q^2 + U^2)^{1/2}$, but is derived from the vector field $\mathbf{P} \equiv (Q, U)$. $|\nabla\mathbf{P}|$ is a gradient in one dimension, for which the appropriate units are (beam)^{−0.5}. Because P measures linearly polarized intensity in units of millijanskys per beam, $|\nabla\mathbf{P}|$ has units of millijanskys per (beam)^{1.5}. The scale showing $|\nabla\mathbf{P}|$ is shown on the right of the image, and ranges from 0 to 15 mJy per (beam)^{1.5}. The inset shows an expanded version of the structure with highest $|\nabla\mathbf{P}|$, covering a box of side 0.9 deg centred on Galactic longitude 329.8 deg and Galactic latitude +1.0 deg. Plotted in the inset is the direction of $\nabla\mathbf{P}$ at each position, defined as

$\arg(\nabla\mathbf{P}) \equiv$

$$\tan^{-1} \left[\text{sign} \left(\frac{\partial Q}{\partial x} \frac{\partial Q}{\partial y} + \frac{\partial U}{\partial x} \frac{\partial U}{\partial y} \right) \sqrt{\left(\frac{\partial Q}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2} / \sqrt{\left(\frac{\partial Q}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial x}\right)^2} \right]$$

For clarity, vectors are shown only at points where the amplitude of the gradient is greater than 5 mJy per (beam)^{1.5}.

frequency dependence of these filaments, because the 1.4-GHz ATCA data shown in Fig. 1 consist of nine independent spectral channels of width 8 MHz, spread over a total bandwidth of 96 MHz. We have constructed images of $|\nabla\mathbf{P}|$ for each individual spectral channel, and these show the same set of specific features as in the overall image, albeit at reduced signal-to-noise ratios. The lack of frequency dependence indicates that the high-gradient structures seen in this data set correspond to physical features in the ISM rather than to contour lines introduced by the particular combination of observing frequency and angular resolution used^{15,17}.

We first consider the possibility that these filaments of high gradient are intrinsic to the source of emission. Abrupt spatial transitions in the strength or geometry of the magnetic field in a synchrotron-emitting region would generate a large gradient in (Q, U) . However, processes of that sort would also produce structure in the overall synchrotron emissivity, such that we would observe features in the image of Stokes I that match those seen in $|\nabla\mathbf{P}|$. No such correspondence is observed, demonstrating that the regions of high polarization gradient are not intrinsic to the source of polarized emission but must be induced by Faraday rotation in magneto-ionized gas.

Because the amount of Faraday rotation is proportional to the line integral of $n_e B_{\parallel}$ from the source to the observer (where n_e is the density of free electrons and B_{\parallel} is the uniform component of the line-of-sight magnetic field), the filamentary structure seen in $|\nabla\mathbf{P}|$ must correspond to boundaries across which n_e and/or B_{\parallel} show a sudden increase or decrease over a small spatial interval. Such discontinuities could be shock fronts or ionization fronts from discrete sources, as have been observed in polarization around the rims of supernova remnants, H II regions and planetary nebulae^{11,18}. We have examined this possibility by carefully comparing our image of $|\nabla\mathbf{P}|$ with images and gradient images of Stokes I (tracing shock waves seen in synchrotron emission; ref. 11), 21-cm H I emission¹⁶ (tracing atomic hydrogen) and 656.3-nm H α emission^{19,20} (tracing ionized hydrogen) over the same field, but do not find any correspondences.

We conclude that the features seen in $|\nabla\mathbf{P}|$ are a generic component of diffuse, ionized gas in this direction in the sky. To test this hypothesis, we performed a series of three-dimensional isothermal simulations of magnetohydrodynamic turbulence in the ISM, each with different parameters for the sonic Mach number, defined as $M_s \equiv \langle |\mathbf{v}| \rangle / c_s$, where \mathbf{v} is the local velocity, c_s is the sound speed and the averaging (indicated by angle brackets) is done over the whole simulation. For each simulation, we propagated a uniform source of polarized emission through the distribution of turbulent, magnetized gas. The resultant Faraday rotation produces a complicated distribution on the sky of Stokes Q and U , from which we generated a map of the polarization gradient using equation (1). Images of $|\nabla\mathbf{P}|$ for representative simulations of the subsonic, transonic and supersonic regimes are shown in Fig. 3. Narrow, elongated filaments of high polarization gradient are apparent in each simulation in Fig. 3, although they differ in their morphology and degree of organization. In particular, the supersonic case (Fig. 3c) shows localized groupings of very high-gradient filaments, corresponding to ensembles of intersecting shocks^{5,21,22}. By contrast, the subsonic (Fig. 3a) and transonic (Fig. 3b) cases show more-diffuse networks of filaments, representing the cusps and discontinuities characteristic of any turbulent velocity field^{6,21,23}.

Visual comparison of the simulated distributions of $|\nabla\mathbf{P}|$ with real data (Fig. 2) suggests that the subsonic and transonic cases shown in Fig. 3a, b more closely resemble the observations than does the supersonic case. We can quantify this statement by calculating the third-order moment (skew, γ) and the fourth-order moment (kurtosis, β) of the probability distribution function of $|\nabla\mathbf{P}|$ for both observations and simulations: these quantities parameterize the degree of Gaussian asymmetry in the probability distribution function, and hence provide information on the amount of compression due to shocks in the data^{6,24}.

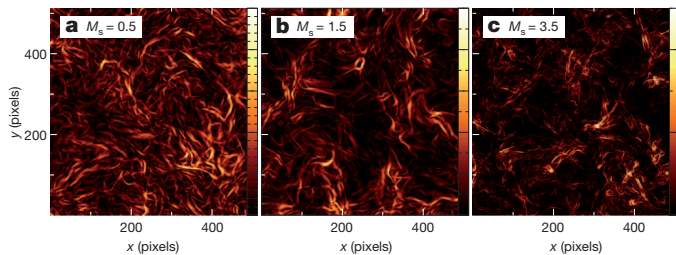


Figure 3 $|\nabla P|$ derived from propagation of linear radio polarization through three different isothermal simulations of magnetized turbulence. Each simulation is a $512 \times 512 \times 512$ -element periodic box with a linear dimension of 0.15 pc for each pixel, evolved in time using an essentially non-oscillatory scheme^{24,30}. Three such simulations are shown, each labelled with its corresponding value for M_s : subsonic ($M_s < 1$; **a**), transonic ($M_s \approx 1$; **b**) and supersonic ($M_s > 1$; **c**). At the start of each simulation, the electron density had a uniform value $n_e = 0.1 \text{ cm}^{-3}$ and the magnetic field was oriented in the plane of the sky with a uniform amplitude of $B = 0.3 \text{ } \mu\text{G}$ (subsonic), $1 \text{ } \mu\text{G}$ (transonic) or $2 \text{ } \mu\text{G}$ (supersonic), which corresponds to a constant Alfvénic Mach number of $M_A = 2$ in each case. Turbulence was driven solenoidally in Fourier space at large scales (small wavenumber) until the turbulent cascade had fully developed and a steady state between input energy and dissipation had been reached. In each case, we illuminated the simulation volume with a background radio source of uniform polarization at an emission frequency of 1.4 GHz, with $Q = 100 \text{ mJy}$ per pixel and $U = 0 \text{ mJy}$ per pixel at all positions. At each pixel, the line integral of $n_e B_{\parallel}$ was computed, and the corresponding Faraday rotation was applied to the polarized signal, to calculate values of Q and U . No effects due to finite angular resolution, depolarization or incomplete interferometric visibility coverage were included, so the observed polarized signal is $P = 100 \text{ mJy}$ per pixel at all positions. We then calculated the gradient, $|\nabla P|$, using equation (1). The scales showing $|\nabla P|$ are shown on the right of the images, and range from 0 to 25 mJy per (pixel)^{1.5} for **a** (interval, 1 mJy per (pixel)^{1.5}), 0 to 100 mJy per (pixel)^{1.5} for **b** (interval, 10 mJy per (pixel)^{1.5}) and 0 to 500 mJy per (pixel)^{1.5} for **c** (interval, 100 mJy per (pixel)^{1.5}).

In the simulations, we found that both the skew and the kurtosis of $|\nabla P|$ increased monotonically with sonic Mach number. We used a genetic algorithm²⁵ to determine that the threshold for strongly supersonic turbulence was $\gamma > 1.1$ and $\beta > 1.5$. We then computed the third- and fourth-order moments for the observed distribution of $|\nabla P|$ shown in Fig. 2, and found that $\gamma = 0.3$ and $\beta = 0.9$.

This analysis of the moments of the polarization gradient therefore confirms quantitatively what we concluded above from visual inspection: the turbulent, ionized ISM in this direction in the sky is subsonic or transonic. The findings we obtained by imaging the polarization gradients produced by interstellar turbulence are supported by recent statistical studies of H α emission measures and of 21-cm H I column densities over large volumes, which have similarly found that $M_s \lesssim 2$ for warm gas throughout the ISM^{26,27}.

In the simulations shown in Fig. 3, the sharp gradients in (Q, U) occur as a result of localized high values of the gas density and magnetic field, resulting from vorticity or shock compression. However, the filamentary features seen in $|\nabla P|$ may not be easily observable in other types of data: for example, if we adopt typical parameters for warm, ionized gas^{1,28} of $n_e \approx 0.3 \text{ cm}^{-3}$ and $B_{\parallel} \approx 2 \text{ } \mu\text{G}$, even the compression associated with a strong adiabatic shock produces across-filament changes in emission measure and Faraday rotation measure of only $\sim 0.5 \text{ pc cm}^{-6}$ and $\lesssim 5 \text{ rad m}^{-2}$, respectively, assuming a spatial scale^{7,14} for these structures of $\sim 0.5 \text{ pc}$. This is below observable levels in H α and other tracers of emission measure. The rotation measure gradient¹⁴ across these interfaces is potentially observable in spectro-polarimetric radio data, but the addition of single-dish observations is required to recover the total power of the polarized signal. By contrast, even a small gradient in rotation measure can produce an arbitrarily large value of $|\nabla P|$ (irrespective of whether single-dish measurements are present in the data), provided that there is a strong source of background polarized emission through which the discontinuities in Faraday rotation are viewed. Further investigation of the polarization

gradient and its statistical properties will provide robust estimates of poorly constrained parameters of turbulent flows such as the sonic and Alfvénic Mach numbers, the characteristic magnetic field strength, the Reynolds number and the physical scale of energy injection.

Received 3 March; accepted 11 August 2011.

Published online 5 October 2011.

1. Ferrière, K. M. The interstellar environment of our galaxy. *Rev. Mod. Phys.* **73**, 1031–1066 (2001).
2. de Avillez, M. A. & Breitschwerdt, D. Global dynamical evolution of the ISM in star forming galaxies. I. High resolution 3D simulations: effect of the magnetic field. *Astron. Astrophys.* **436**, 585–600 (2005).
3. Chepurnov, A. & Lazarian, A. Extending the big power law in the sky with turbulence spectra from Wisconsin H α Mapper data. *Astrophys. J.* **710**, 853–858 (2010).
4. Armstrong, J. W., Rickett, B. J. & Spangler, S. R. Electron density power spectrum in the local interstellar medium. *Astrophys. J.* **443**, 209–221 (1995).
5. Kritsuk, A. G., Norman, M. L., Padoan, P. & Wagner, R. The statistics of supersonic isothermal turbulence. *Astrophys. J.* **665**, 416–431 (2007).
6. Kowal, G., Lazarian, A. & Beresnyak, A. Density fluctuations in MHD turbulence: Spectra, intermittency, and topology. *Astrophys. J.* **658**, 423–445 (2007).
7. Kissmann, R., Kleimann, J., Fichtner, H. & Grauer, R. Local turbulence simulations for the multiphase ISM. *Mon. Not. R. Astron. Soc.* **391**, 1577–1588 (2008).
8. Wieringa, M. H., de Bruyn, A. G., Jansen, D., Broun, W. N. & Katgert, P. Small scale polarization structure in the diffuse galactic emission at 325 MHz. *Astron. Astrophys.* **268**, 215–229 (1993).
9. Gray, A. D., Landecker, T. L., Dewdney, P. E. & Taylor, A. R. A large-scale, interstellar Faraday-rotation feature of unknown origin. *Nature* **393**, 660–662 (1998).
10. Haverkorn, M., Katgert, P. & de Bruyn, A. G. Structure in the local Galactic ISM on scales down to 1 pc, from multi-band radio polarization observations. *Astron. Astrophys.* **356**, L13–L16 (2000).
11. Gaensler, B. M. *et al.* Radio polarization from the inner galaxy at arcminute resolution. *Astrophys. J.* **549**, 959–978 (2001).
12. Uyaniker, B. & Landecker, T. L. A highly ordered Faraday rotation structure in the interstellar medium. *Astrophys. J.* **575**, 225–233 (2002).
13. Shukurov, A. & Berkhuijsen, E. M. Faraday ghosts: depolarization canals in the galactic radio emission. *Mon. Not. R. Astron. Soc.* **342**, 496–500 (2003).
14. Haverkorn, M. & Heitsch, F. Canals beyond Mars: beam depolarization in radio continuum maps of the warm ISM. *Astron. Astrophys.* **421**, 1011–1019 (2004).
15. Fletcher, A. & Shukurov, A. Depolarization canals and interstellar turbulence. *EAS Publ. Ser.* **23**, 109–128 (2007).
16. McClure-Griffiths, N. M. *et al.* The Southern Galactic Plane Survey: the test region. *Astrophys. J.* **551**, 394–412 (2001).
17. Newton-McGee, K. J. *Radio Polarimetry as a Probe of Interstellar Magnetism* 83–87. Ph.D. thesis, Univ. Sydney (2009).
18. Ransom, R. R., Uyaniker, B., Kothes, R. & Landecker, T. L. Probing the magnetized interstellar medium surrounding the planetary nebula Sh 2–216. *Astrophys. J.* **684**, 1009–1017 (2008).
19. Gaustad, J. E., McCullough, P. R., Rosing, W. & Van Buren, D. A robotic wide-angle H α survey of the southern sky. *Publ. Astron. Soc. Pacif.* **113**, 1326–1348 (2001).
20. Parker, Q. A. *et al.* The AAO/UKST SuperCOSMOS H α survey. *Mon. Not. R. Astron. Soc.* **362**, 689–710 (2005).
21. Beresnyak, A., Lazarian, A. & Cho, J. Density scaling and anisotropy in supersonic magnetohydrodynamic turbulence. *Astrophys. J.* **624**, L93–L96 (2005).
22. Lemaster, M. N. & Stone, J. M. Dissipation and heating in supersonic hydrodynamic and MHD turbulence. *Astrophys. J.* **691**, 1092–1108 (2009).
23. Eyink, G. L. Besov spaces and the multifractal hypothesis. *J. Stat. Phys.* **78**, 353–375 (1995).
24. Burkhart, B., Falceta-Gonçalves, D., Kowal, G. & Lazarian, A. Density studies of MHD interstellar turbulence: statistical moments, correlations and bispectrum. *Astrophys. J.* **693**, 250–266 (2009).
25. Whitley, D. A genetic algorithm tutorial. *Stat. Comput.* **4**, 65–85 (1994).
26. Hill, A. S. *et al.* The turbulent warm ionized medium: emission measure distribution and MHD simulations. *Astrophys. J.* **686**, 363–378 (2008).
27. Burkhart, B., Stanimirović, S., Lazarian, A. & Kowal, G. Characterizing magnetohydrodynamic turbulence in the Small Magellanic Cloud. *Astrophys. J.* **708**, 1204–1220 (2010).
28. Gaensler, B. M., Madsen, G. J., Chatterjee, S. & Mao, S. A. The vertical structure of warm ionised gas in the Milky Way. *Publ. Astron. Soc. Aust.* **25**, 184–200 (2008).
29. Haverkorn, M., Gaensler, B. M., McClure-Griffiths, N. M., Dickey, J. M. & Green, A. J. The Southern Galactic Plane Survey: polarized radio continuum observations and analysis. *Astrophys. J.* **167** (suppl.), 230–238 (2006).
30. Cho, J. & Lazarian, A. Compressible magnetohydrodynamic turbulence: mode coupling, scaling relations, anisotropy, viscosity-damped regime and astrophysical implications. *Mon. Not. R. Astron. Soc.* **345**, 325–339 (2003).

Acknowledgements We thank S. Brown, A. Hill, R. Kissmann, A. MacFadyen, M.-M. Mac Low, E. Petroff, P. Plane and X. Sun for discussions. The Australia Telescope Compact Array is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. B.M.G. and T.R. acknowledge the support of the Australian Research Council through grants FF0561298, FL100100114 and FS100100033. B.B. acknowledges support from the National Science Foundation Graduate Research Fellowship and the NASA Wisconsin Space Grant Institution. A.L. acknowledges the support of the National Science Foundation through grant AST0808118 and of the

Center for Magnetic Self-Organization in Astrophysical and Laboratory Plasmas. We thank the staff of the Australia Telescope National Facility, especially M. Calabretta, R. Haynes, D. McConnell, J. Reynolds, R. Sault, R. Wark and M. Wieringa, for their support of the Southern Galactic Plane Survey.

Author Contributions J.M.D., N.M.Mc.-G., B.M.G. and A.J.G. carried out the original observations. B.M.G., N.M.Mc.-G. and T.R. produced the polarization images from the raw data. B.M.G., M.H., K.J.N.-Mc., R.D.E and N.M.Mc.-G. worked together to develop the gradient technique, and B.M.G. then applied the gradient technique to the images. B.B.

and A.L. performed the simulations and the statistical analysis. B.M.G. led the writing of the paper and the interpretation of results. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.M.G. (bryan.gaensler@sydney.edu.au).

Ocean-like water in the Jupiter-family comet 103P/Hartley 2

Paul Hartogh¹, Dariusz C. Lis², Dominique Bockelée-Morvan³, Miguel de Val-Borro¹, Nicolas Biver³, Michael Küppers⁴, Martin Emprechtinger², Edwin A. Bergin⁵, Jacques Crovisier³, Miriam Rengel¹, Raphael Moreno³, Sławomira Szutowicz⁶ & Geoffrey A. Blake²

For decades, the source of Earth's volatiles, especially water with a deuterium-to-hydrogen ratio (D/H) of $(1.558 \pm 0.001) \times 10^{-4}$, has been a subject of debate. The similarity of Earth's bulk composition to that of meteorites known as enstatite chondrites¹ suggests a dry proto-Earth² with subsequent delivery of volatiles³ by local accretion⁴ or impacts of asteroids or comets^{5,6}. Previous measurements in six comets from the Oort cloud yielded a mean D/H ratio of $(2.96 \pm 0.25) \times 10^{-4}$. The D/H value in carbonaceous chondrites, $(1.4 \pm 0.1) \times 10^{-4}$, together with dynamical simulations, led to models in which asteroids were the main source of Earth's water⁷, with ≤ 10 per cent being delivered by comets. Here we report that the D/H ratio in the Jupiter-family comet 103P/Hartley 2, which originated in the Kuiper belt, is $(1.61 \pm 0.24) \times 10^{-4}$. This result substantially expands the reservoir of Earth ocean-like water to include some comets, and is consistent with the emerging picture of a complex dynamical evolution of the early Solar System^{8,9}.

On 17 November 2010, using the Herschel Space Observatory, we determined the D/H ratio in a comet from a reservoir other than the Oort cloud—103P/Hartley 2. Such Jupiter-family comets are believed to originate from the Kuiper belt, which exists beyond the orbits of the giant planets at radii between 30 and 50 astronomical units¹⁰ (1 AU is the average Earth–Sun distance). In contrast, Oort-cloud comets are theorized to have originated from radii near the gas giants and to have been subsequently ejected to the Oort cloud ($>5,000$ AU)¹¹. The Herschel measurement therefore traces the water D/H ratio in a new population of water-ice-rich bodies in the Solar System that are a potential source of water on the Earth.

To obtain an accurate determination of the D/H ratio in water, we carried out simultaneous observations of optically thin isotopic variants of water, specifically HDO and H₂¹⁸O (Fig. 1), as part of our Solar System observing programme¹². This was critical for comet 103P/Hartley 2, whose activity and water outgassing rates exhibited significant short-term variations¹³. We used state-of-the-art excitation models to determine the HDO and H₂¹⁸O beam integrated column densities and production rates from the measured line intensities. Observation and modelling details are given in Supplementary Information. A critical point is that all observations sampled the same region of the coma, about 6,500 km in diameter.

The retrieved gas column densities and production rates are sensitive to collisional cross-sections, along with the density and temperature profiles of H₂O and electrons, and we thus considered a range of model parameters (Table 1). Although the production rates determined for the various model parameters differ slightly, the value of the D/H ratio is estimated to be $(1.61 \pm 0.24) \times 10^{-4}$. In our analysis, we assumed an H₂¹⁶O/H₂¹⁸O ratio of 500 ± 50 , a range that encompasses the Earth value and is consistent with previous measurements in cometary water¹⁴ (see also Supplementary Information). The quoted 1σ uncertainty in the D/H ratio includes a 5% uncertainty related to modelling.

Our measured D/H value is substantially larger than that which characterized the young Sun (4.5 Gyr ago; the protosolar ratio), believed to be about 2.1×10^{-5} , which in turn is slightly higher than the value found in the local interstellar medium today (1.6×10^{-5}) and

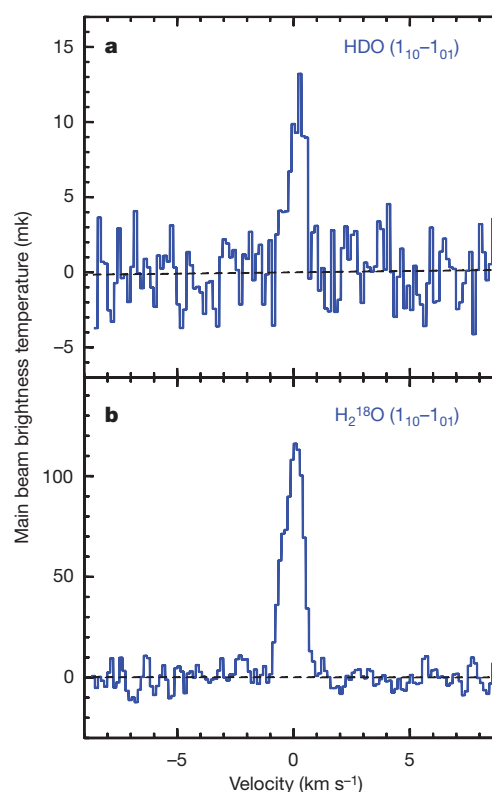


Figure 1 | Submillimetre water emission lines from comet 103P/Hartley 2. The time of the observations was 20 days after perihelion, when the comet was 1.095 AU from the Sun and 0.212 AU from Herschel. Because the H₂O ground state rotational lines in comets are optically thick^{29,30}, observations of the rare oxygen isotopic counterpart, H₂¹⁸O, provide a more reliable reference for the D/H determination. The spectra of the $1_{10}-1_{01}$ lines of HDO (a) and H₂¹⁸O (b) at 509.292 and 547.676 GHz, respectively, were obtained with the Heterodyne Instrument for the Far Infrared (HIFI) High Resolution Spectrometer (HRS) between 17.28 and 17.64 November 2010 UT. The line intensities, expressed in the main-beam brightness temperature scale, are 0.011 ± 0.001 and 0.117 ± 0.002 K km s⁻¹, for HDO and H₂¹⁸O respectively, averaging the two instrument polarizations. The velocity scale is given relative to the velocity of the comet's nucleus. The spectral resolution is 141 and 132 m s⁻¹ for the HDO and H₂¹⁸O spectra, respectively. For details of the observational sequence and basic parameters of the data analysis, see Supplementary information.

¹Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany. ²California Institute of Technology, Pasadena, California 91125, USA. ³LESIA-Observatoire de Paris, CNRS, UPMC, Université Paris-Diderot, 5 place Jules Janssen, 92195 Meudon, France. ⁴Rosetta Science Operations Centre, European Space Astronomy Centre, 28691 Villanueva de la Cañada, Madrid, Spain. ⁵Astronomy Department, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁶Space Research Centre, Polish Academy of Sciences, 00-716 Warsaw, Poland.

Table 1 | Calculating the D/H ratio in water in comet 103P/Hartley 2

Model	T_{gas} (K)	x_{ne}	$\langle N(\text{HDO}) \rangle$ (10^{10} cm^{-2})	$Q(\text{HDO})$ (10^{24} s^{-1})	$\langle N(\text{H}_2^{18}\text{O}) \rangle$ (10^{11} cm^{-2})	$Q(\text{H}_2^{18}\text{O})$ (10^{25} s^{-1})	D/H
(1)	50	0.	4.9	3.1	3.5	2.1	1.49×10^{-4}
	50	0.2	3.6	2.3	2.5	1.5	1.55×10^{-4}
	70	0.2	3.7	2.4	2.5	1.5	1.60×10^{-4}
	Law	0.1	5.7	3.6	3.7	2.2	1.63×10^{-4}
(2)	50	0.2	4.3	2.7	2.9	1.8	1.54×10^{-4}
	Law	0.1	4.8	3.1	3.2	1.9	1.58×10^{-4}

The parameter x_{ne} , scaling the electron density profile in the models, is constrained by mapping observations. $\langle N \rangle$ and Q are respectively beam integrated column density and production rate, determined using different parameters in the excitation models^{14,29}. Production rates were computed assuming isotropic outflow of water from the nucleus, with a velocity of 0.6 km s^{-1} , consistent with the width of the H_2^{18}O line. We accounted for the $10''$ offset between the centre of the beam and the position of the peak of the H_2O distribution. Values of 50 and 70 K for the gas kinetic temperature, T_{gas} , are consistent with multi-transition measurements of gaseous species in the millimetre and near-infrared range, respectively. The gas kinetic temperature is expected to decrease with increasing distance owing to quasi-adiabatic expansion of the escaping gases: the temperature law assumes that $T_{\text{gas}} = 80 \text{ K}$ for $r < 270 \text{ km}$, $T_{\text{gas}} = 12 \text{ K}$ for $r > 630 \text{ km}$, with a linear decrease between 270 and 630 km, where r is the distance from the nucleus. Collision cross-sections involving water molecules and electrons are modelled differently in models (1) and (2). Both models use an electron density profile based on *in situ* measurements in comet 1P/Halley scaled to the activity of 103P/Hartley 2 (ref. 29). The D/H ratio is equal to $0.5 \times Q(\text{HDO})/Q(\text{H}_2\text{O})$, with $Q(\text{H}_2\text{O}) = 500 \times Q(\text{H}_2^{18}\text{O})$. See Supplementary information for details of the models and model parameters.

comparable to the primordial D/H ratio in the Universe after the Big Bang (Fig. 2)¹⁵. Protosolar water, on the other hand, is believed to be highly enriched ($\text{D/H} \approx 1 \times 10^{-3}$)¹⁶ due to the low-temperature (~ 10 – 30 K) non-equilibrium chemistry that characterizes the dense interstellar medium¹⁷, either via gas-phase isotopic exchange reactions involving ions and radicals, or grain-surface processes. Consequently, the resulting D/H ratio in water ice is very sensitive to the physical conditions, in particular the kinetic temperature of the medium. After the protosolar cloud collapsed to form the solar nebula, isotopic exchange reactions between molecular hydrogen and HDO molecules would have led to a gradual reduction of D/H in water¹⁸, as compared to the initial interstellar medium value. Because the efficiency of these reactions and the turbulent mixing within the solar nebula is correlated with the gas density and temperature, the deuterium enhancement in water has been predicted to increase with the heliocentric distance^{19–21}. Ices, captured by planetesimals and cometesimals, would have then preserved the deuterium enrichment in water from this early epoch. As

a result, small Solar System bodies are expected to exhibit different D/H ratios in their water ice depending on the distance from the Sun at which they were formed.

In the context of this simple nebular model, the D/H ratio of $(1.61 \pm 0.24) \times 10^{-4}$ in comet 103P/Hartley 2—a factor of two lower than that measured in Oort-cloud comets (Fig. 2) and, within uncertainties, consistent with that of the Earth's oceans (for which the Vienna Standard Mean Ocean Water (VSMOW) value is $(1.558 \pm 0.001) \times 10^{-4}$)—is therefore surprising, and compatible with two different schemes: (1) either this comet did not form in a region that was further from the Sun than the assembly zone of the Oort-cloud comets, or (2) the dependence of the water D/H ratio with distance from the Sun is not as expected on the basis of current models. Concerning the first possibility, dynamical models indeed suggest that a fraction of the Jupiter-family comets originate in the Oort cloud²². Still, even if comet 103P/Hartley 2 stems from the Oort cloud, this would not explain why its D/H ratio is different from that seen in other Oort-cloud comets. Models also suggest that a fraction of the Jupiter-family comets may have originated from the Trojan asteroid swarms sharing the orbit of Jupiter²³. The Trojans are generally thought to have resided at their current location since the formation of the Solar System. Therefore, Jupiter family comets originating in the Trojan region could, in theory, display deuterium enrichment values lower than those for bodies originating in the Kuiper belt, if they indeed formed in the vicinity of Jupiter. However, the most probable scenario is that 103P/Hartley 2 originated in the Kuiper belt.

It is difficult to explain the low D/H ratio in 103P/Hartley 2 (compared to that of previous measurements in comets) with the formation regions of comets, thus models of the gradient of D/H in the Solar System—predictions not yet directly confirmed by observations, owing to scarcity of accurate isotopic measurements—may need to be revisited. In fact, one recent model has suggested that the D/H ratio of water vapour can be locally enhanced²⁴. However, the vapour must then be implanted into cometary ices. Moreover, until the measurement of 103P/Hartley 2 there was no observational confirmation of variations in the D/H ratio. One possible solution is that there was large-scale movement of material between the inner and outer Solar

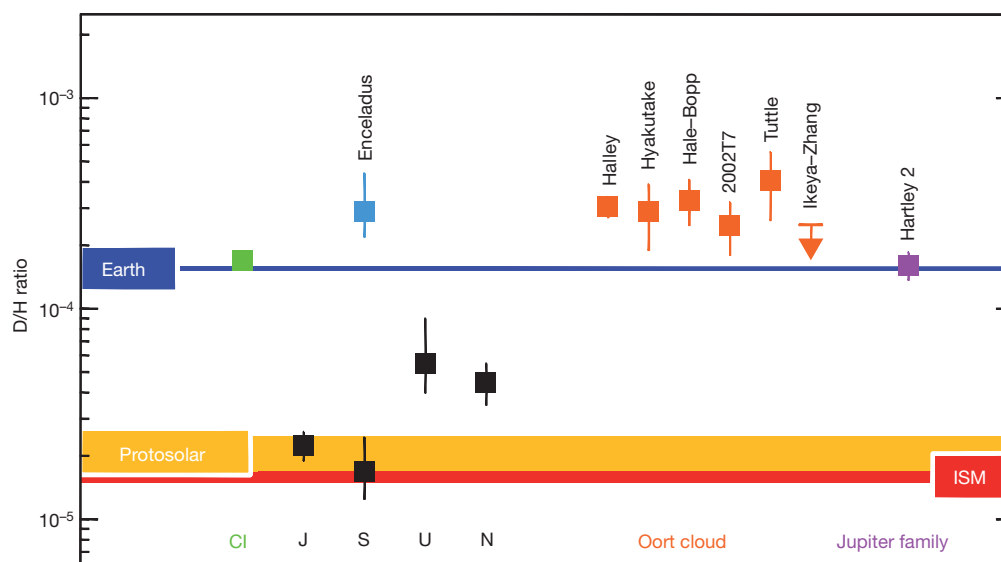


Figure 2 | D/H ratios in the Solar System. Orange squares, values measured for water in the Oort-cloud comets 1P/Halley, C/1996 B2 (Hyakutake), C/1995 O1 (Hale-Bopp), C/2002 T7 (LINEAR) and 8P/Tuttle. Arrow (for 153P/Ikeya-Zhang), upper limit. Purple square, present measurement in the water of 103P/Hartley 2. Black symbols, D/H ratio in H_2 in the atmosphere of the giant planets—Jupiter (J), Saturn (S), Uranus (U) and Neptune (N). Light blue and green symbols, D/H values for water in the plume of Saturn's moon Enceladus and in CI carbonaceous chondrites, respectively. Error bars, 1σ . The D/H

determinations in comets originating from the Oort cloud are twice the value for the Earth's ocean (blue line) and about a factor of ten larger than the protosolar value in H_2 (broad yellow line), the latter being comparable to the value in atomic hydrogen found in the local interstellar medium (ISM, red horizontal line). The D/H ratio in the Jupiter-family comet 103P/Hartley 2 is the same as the Earth's ocean value and the chondritic CI value. Uranus and Neptune have been enriched in deuterium by the mixing of their atmospheres with D-rich protoplanetary ices. For further details, see Supplementary Table 1.

System. According to a recent theory proposed for the early Solar System (the Grand Tack scenario), when the giant planets were still embedded in the nebular gas disk, there was a general radial mixing of the distribution of comets and asteroids born in different regions²⁵. The similarity of the D/H ratio in comet 103P/Hartley 2, which probes the Kuiper belt, with that found in CI chondrites tracing the asteroid belt, would be in agreement with a general shake-up of the Solar System at early times.

In more 'orderly' models, the high D/H values derived from the earlier observations of Oort-cloud comets suggested that at most 10% of the Earth's water could have been supplied from the outermost Solar System, but even under these circumstances a number of scenarios have been developed suggesting that terrestrial water could have in fact been delivered by comets. Such models are based on assumptions about the heliocentric D/H gradient²⁶, and the analysis of lunar samples²⁷ and telluric sedimentary rocks formed at the end of the Late Heavy Bombardment phase²⁸. Our Herschel observations of a VSMOW-like D/H ratio in 103P/Hartley 2 enlarge the region of the solar nebula known to have a D/H ratio similar to that of Earth's oceans; this region now includes both the asteroid belt and the much larger Kuiper belt, thereby providing support for the theory of a common water source for the inner Solar System bodies (including the Earth) in which comets play an important part.

Further constraints on the delivery of volatiles to the early Earth and an improved understanding of the origin of the different dynamical classes of comets will require significantly larger sample sizes than those at present available. A handful of additional measurements can be expected from Herschel before its cryogen supply is exhausted, but the comparison of D/H ratios in the inner and outer Solar System must necessarily utilize very different objects and materials. For the inner Solar System, *in situ* space missions or sample return missions to the outer asteroid belt would provide critical new data. Astronomically, the minuscule strength of HDO spectroscopic signatures makes D/H measurements extremely challenging, and dedicated programmes using new facilities will be required to substantially increase the inventory of high-precision D/H ratios in comets and other icy Solar System bodies, including the Jovian satellites.

Received 16 May; accepted 30 August 2011.

Published online 5 October 2011.

1. Javoy, M. *et al.* The chemical composition of the Earth: enstatite chondrite models. *Earth Planet. Sci. Lett.* **293**, 259–268 (2010).
2. Wänke, H. Constitution of terrestrial planets. *Phil. Trans. R. Soc. Lond.* **303**, 287–302 (1981).
3. Robert, F. A distinct source of lunar water? *Nature Geosci.* **4**, 74–75 (2011).
4. Drake, M. J. & Righter, K. Determining the composition of the Earth. *Nature* **416**, 39–44 (2002).
5. Oró, J. Comets and the formation of biochemical compounds on the primitive Earth. *Nature* **190**, 389–390 (1961).
6. Owen, T., Bar-Nun, A. & Kleinfeld, I. Possible cometary origin of heavy noble gases in the atmospheres of Venus, Earth and Mars. *Nature* **358**, 43–46 (1992).
7. Morbidelli, A. *et al.* Source regions and time scales for the delivery of water to Earth. *Meteorit. Planet. Sci.* **35**, 1309–1320 (2000).
8. Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435**, 459–461 (2005).
9. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469 (2005).
10. Levison, H. F. & Duncan, M. J. From the Kuiper belt to Jupiter-family comets: the spatial distribution of ecliptic comets. *Icarus* **127**, 13–32 (1997).
11. Dones, L., Weissman, P. R., Levison, H. F. & Duncan, M. J. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 153–174 (Univ. Arizona Press, 2005).
12. Hartogh, P. *et al.* Water and related chemistry in the solar system: a guaranteed time key programme for Herschel. *Planet. Space Sci.* **57**, 1596–1606 (2009).

13. Meech, K. *et al.* EPOXI: 103P/Hartley 2 observations from a worldwide campaign. *Astrophys. J.* **734**, L1 (2011).
14. Biver, N. *et al.* Submillimetre observations of comets with Odin: 2001–2005. *Planet. Space Sci.* **55**, 1058–1068 (2007).
15. Linsky, J. L. *et al.* What is the total deuterium abundance in the local galactic disk? *Astrophys. J.* **647**, 1106–1124 (2006).
16. Butner, H. M. *et al.* Discovery of interstellar heavy water. *Astrophys. J.* **659**, L137–L140 (2007).
17. Watson, W. D. Ion-molecule reactions, molecule formation, and hydrogen-isotope exchange in dense interstellar clouds. *Astrophys. J.* **188**, 35–42 (1974).
18. Geiss, J. & Reeves, H. Cosmic and solar system abundances of deuterium and helium-3. *Astron. Astrophys.* **18**, 126–132 (1972).
19. Drouart, A., Dubrulle, B., Gautier, D. & Robert, F. Structure and transport in the solar nebula from constraints on deuterium enrichment and giant planet formation. *Icarus* **140**, 129–155 (1999).
20. Mousis, O. *et al.* Constraints on the formation of comets from D/H ratios measured in H₂O and HCN. *Icarus* **148**, 513–525 (2000).
21. Aikawa, Y. & Herbst, E. Two-dimensional distributions and column densities of gaseous models in protoplanetary disks II. Deuterated species and UV shielding by ambient clouds. *Astron. Astrophys.* **371**, 1107–1117 (2001).
22. Emel'yanenko, V. V., Asher, D. J. & Bailey, M. E. Centaurs from the Oort cloud and the origin of Jupiter-family comets. *Mon. Not. R. Astron. Soc.* **361**, 1345–1351 (2005).
23. Marzari, F., Farinella, P. & Vanzani, V. Are Trojan collisional families a source for short-period comets? *Astron. Astrophys.* **299**, 267–276 (1995).
24. Thi, W.-F., Woitke, P. & Kamp, I. Warm non-equilibrium gas phase chemistry as a possible origin of high HDO/H₂O ratios in hot and dense gases: application to inner protoplanetary discs. *Mon. Not. R. Astron. Soc.* **407**, 232–246 (2010).
25. Walsh, K. J. & Morbidelli, A. The effect of an early planetesimal-driven migration of the giant planets on terrestrial planet formation. *Astron. Astrophys.* **526**, A126 (2011).
26. Delsemme, A. H. The deuterium enrichment observed in recent comets is consistent with the cometary origin of seawater. *Planet. Space Sci.* **47**, 125–131 (1998).
27. Greenwood, J. P. *et al.* Hydrogen isotope ratios in lunar rocks indicate delivery of cometary water to the Moon. *Nature Geosci.* **4**, 79–82 (2011).
28. Jørgensen, U. G. *et al.* The Earth-Moon system during the late heavy bombardment period — geochemical support for impacts dominated by comets. *Icarus* **204**, 368–380 (2009).
29. Zakharov, V., Bockelée-Morvan, D., Biver, N., Crovisier, J. & Lecacheux, A. Radiative transfer simulation of water rotational excitation in comets. Comparison of the Monte Carlo and escape probability methods. *Astron. Astrophys.* **473**, 303–310 (2007).
30. Bensch, F. & Bergin, E. A. The pure rotational line emission of ortho-water in comets. I. Radiative transfer model. *Astrophys. J.* **615**, 531–544 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation by NASA. The Heterodyne Instrument for the Far Infrared (HIFI) has been designed and built by a consortium of institutes and university departments from across Europe, Canada and the United States under the leadership of SRON, the Netherlands Institute for Space Research, and with major contributions from Germany, France and the USA. This development has been supported by national funding agencies: CEA, CNES, CNRS (France); ASI (Italy); and DLR (Germany). Additional funding support for some instrument activities has been provided by ESA. Support for this work was also provided by NASA through an award issued by JPL/Caltech. D.C.L. is supported by a NSF award to the Caltech Submillimeter Observatory. We thank R. Lorente, P. García-Lario, M. Kidger and G. Pilbratt for helping with the scheduling of these observations, and I. Avruch for the assistance with HIFI specific software issues.

Author Contributions This paper represents the combined work of the HsO (the Herschel guaranteed time key programme "Water and related chemistry in the solar system") team members listed as authors. P.H. is the coordinator of this programme. All authors contributed to this work, including observation planning, data analysis and writing of the manuscript. N.B., D.B.-M., M.R., R.M., M.d.V.-B. and M.E. carried out the data reduction and contributed to the modelling efforts. All authors were collectively involved in the discussion and interpretation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.H. (hartogh@mps.mpg.de).

Coherent coupling of a superconducting flux qubit to an electron spin ensemble in diamond

Xiaobo Zhu¹, Shiro Saito¹, Alexander Kemp¹, Kosuke Kakuyanagi¹, Shin-ichi Karimoto¹, Hayato Nakano¹, William J. Munro¹, Yasuhiro Tokura¹, Mark S. Everitt², Kae Nemoto², Makoto Kasu¹, Norikazu Mizuochi^{3,4} & Kouichi Semba¹

During the past decade, research into superconducting quantum bits (qubits) based on Josephson junctions has made rapid progress¹. Many foundational experiments have been performed^{2–8}, and superconducting qubits are now considered one of the most promising systems for quantum information processing. However, the experimentally reported coherence times are likely to be insufficient for future large-scale quantum computation. A natural solution to this problem is a dedicated engineered quantum memory based on atomic and molecular systems. The question of whether coherent quantum coupling is possible between such natural systems and a single macroscopic artificial atom has attracted considerable attention^{9–12} since the first demonstration of macroscopic quantum coherence in Josephson junction circuits². Here we report evidence of coherent strong coupling between a single macroscopic superconducting artificial atom (a flux qubit) and an ensemble of electron spins in the form of nitrogen–vacancy colour centres in diamond. Furthermore, we have observed coherent exchange of a single quantum of energy between a flux qubit and a macroscopic ensemble consisting of about 3×10^7 such colour centres. This provides a foundation for future quantum memories and hybrid devices coupling microwave and optical systems.

With the early successes of single-atom quantum state manipulation¹³, research in quantum information processing with atomic systems has largely progressed independently from that using solid-state systems. In recent years, considerable effort has been devoted to coupling atomic and molecular systems to solid-state qubits to form hybrid quantum devices^{9–11}. Hybrid devices involving the integration of an ensemble of atomic spin systems with a transmission line resonator have been realized^{14–18}. Such schemes have the potential to couple superconducting solid-state qubits to optical fields via atomic systems, thus allowing quantum media conversion. The coupling strength, g , of an individual atomic system to one electromagnetic mode in a resonator circuit is usually too small for the coherent exchange of quantum information. However, the coupling strength of an ensemble of N such atomic systems will be enhanced by a factor of \sqrt{N} (ref. 19), making it possible to reach the strong-coupling regime ($g\sqrt{N} \gg \kappa$ and $g\sqrt{N} \gg \gamma$, where κ and γ are the respective damping rates of the resonator circuit and the atomic system).

Of the many possible hybrid systems, coupling a flux qubit to a nitrogen–vacancy colour centre (NV[−] centre) in diamond is particularly appealing. First, the magnetic coupling strength between a flux qubit and a single NV[−] centre can be three orders of magnitude larger than that between a superconducting transmission line resonator and an NV[−] centre¹². Second, the ground state of a NV[−] centre is a triplet ($S = 1$) owing to its C_{3v} symmetry (Fig. 1b). The $S = 1$ state $|m_S = 0\rangle$ is separated by 2.88 GHz from the near-degenerate excited states $|m_S = \pm 1\rangle$ under zero magnetic field (Fig. 1c). This energy separation is ideal for a gap-tunable flux qubit to be brought into and out of resonance with it.

In this Letter, we report the observation of vacuum Rabi oscillations between a flux qubit and an ensemble of approximately 3×10^7 NV[−] centres in diamond. This demonstrates strong coherent coupling between two dissimilar quantum systems with a collective coupling constant of $g_{\text{ens}} \approx 70$ MHz.

We begin by describing our experimental set-up (Fig. 1). A sample of diamond containing NV[−] centres was prepared by ion implantation of $^{12}\text{C}^{2+}$ at 700 keV under high vacuum into single-crystal, type-1b diamond synthesized under high pressure and at high temperature (HPHT) and with (001) surface orientation. The $^{12}\text{C}^{2+}$ ions, implanted with a surface density of $3 \times 10^{13} \text{ cm}^{-2}$, were stopped at a depth of $600^{+50}_{-100} \text{ nm}$. This generated vacancies with a concentration of order $5 \times 10^{18} \text{ cm}^{-3}$ over a depth of $\sim 0.7 \mu\text{m}$. After implantation, the crystals were annealed at 900 °C under vacuum for 3 h. This high-dose carbon implantation method enhances the yield of generated

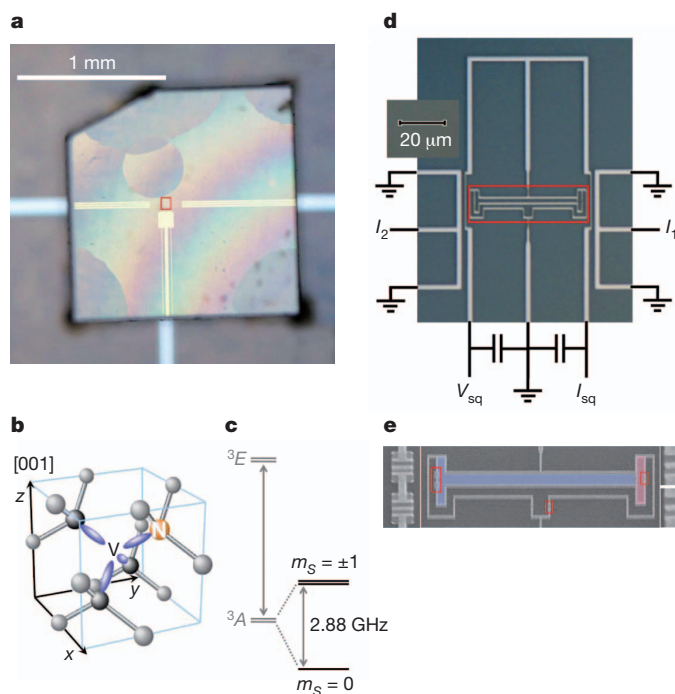


Figure 1 | Experimental set-up of a NV[−] diamond sample attached to a flux qubit system. **a**, Diamond crystal glued on top of a flux qubit (red box). **b**, NV[−] centre: V, vacancy; N, nitrogen atom. **c**, Energy diagram of the NV[−] centre²⁸. **d**, Optical micrograph of the superconducting circuits. The flux qubit and a superconducting quantum interference device (SQUID) detector (current, I_{sq} ; voltage, V_{sq}) are highlighted (red box). Microwave control lines (I_1 and I_2) are located on both sides. **e**, Gap-tunable flux qubit and the edge-shared SQUID used as qubit state detector. The qubit contains two loops, the main loop (blue) and the α -loop (SQUID; pink), which controls the tunnelling energy of the qubit. Josephson junctions are highlighted (red boxes).

¹NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan. ²National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. ³University of Osaka, Graduate school of Engineering Science, 1-3 Machikane-yama, Toyonaka, Osaka 560-8531, Japan. ⁴PRESTO JST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.

NV^- centres²⁰. Photoluminescence optical spectroscopy (Fig. 2) established that NV^- centres were generated with a concentration of $\sim 1.1 \times 10^{18} \text{ cm}^{-3}$ over a depth of $1 \mu\text{m}$ (see Supplementary Information, section II, for further details). We can describe the ground state of a single NV^- centre with the Hamiltonian²¹

$$H_{\text{NV}} = hDS_z^2 + hE(S_x^2 - S_y^2) + hg_e\mu_B\mathbf{B}\cdot\mathbf{S}$$

where S_x , S_y and S_z are the usual Pauli spin-one operators, the components of \mathbf{S} ; D is the zero-field splitting (2.878 GHz); E is the strain-induced splitting ($< 10 \text{ MHz}$); $g_e = 2$ is the nitrogen-vacancy Landé factor; $\mu_B = 14 \text{ MHz mT}^{-1}$ is the Bohr magneton; and h is Planck's constant. The final term represents the Zeeman splitting, which is negligible in our case because the magnetic field (\mathbf{B}) applied perpendicular to the surface of the chip to prepare the flux qubit is less than 0.1 mT .

A diamond crystal was glued on top of the superconducting circuit with the ^{12}C -implanted (001) surface facing the flux qubit (Fig. 1a). The distance between the flux qubit and the surface of the diamond crystal was adjusted to be less than a micrometre. We used a gap-tunable flux qubit^{22,23} (Fig. 1d, e and Supplementary Information, section I) in which the smallest of the three Josephson junctions in the qubit was replaced by a low-inductance d.c. SQUID loop (Fig. 1e, pink loop). The flux qubit/ NV^- ensemble coupled system was measured by detecting the qubit state using a read-out d.c. SQUID (the biggest loop in Fig. 1e sharing edges with the flux qubit), which is inductively coupled to the qubit. When biasing the main loop at close to one-half of a flux quantum, the device is effectively a two-level system²⁴ described by the Hamiltonian

$$H_{\text{qb}} = \frac{h}{2}(\Delta\sigma_x + \varepsilon\sigma_z)$$

which is given in the basis of clockwise and anticlockwise qubit persistent currents. Here σ_x and σ_z are Pauli spin-half operators, $h\varepsilon = 2I_p(\Phi_{\text{ex}} - 3\Phi_0/2)$ is the energy bias ($I_p \approx 300 \text{ nA}$ is the persistent current in the qubit, Φ_{ex} is the external flux threading the qubit loop and $\Phi_0 = h/2e$ is the flux quantum) and Δ is the tunnel splitting. The energy splitting of the gap-tunable flux qubit is $hF = h\sqrt{\varepsilon^2 + \Delta^2}$ where ε and Δ can be controlled independently using the external magnetic flux threading the two loops. This type of flux qubit can be tuned into resonance with a NV^- ensemble *in situ* at a base temperature of

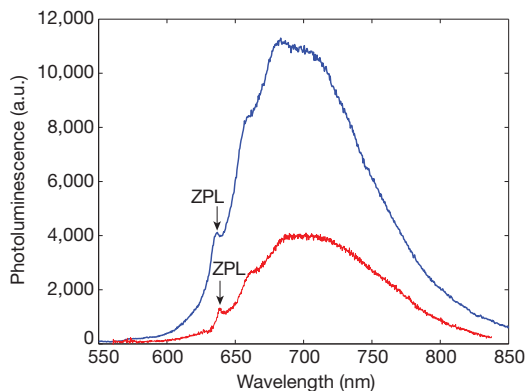


Figure 2 | Photoluminescence spectra. Spectra of ensembles of colour centres in the highly carbon-implanted sample (blue) and a single NV^- centre in pure diamond at room temperature (red). The signal intensities are normalized for comparison (a.u., arbitrary units). In the spectrum of the single NV^- centre, the contributions of phonon Raman scattering from diamond at 573 nm and between 600 and 620 nm have been subtracted²⁹. The NV^- zero-phonon line (ZPL) at 637 nm is observed in both spectra^{29,30}. The spectra are similar to each other and those reported previously^{29,30}, indicating that NV^- centres are abundant in the implanted sample. The signal intensity of the ensemble is $\sim 6.5 \times 10^4$ that of the single centre, and we estimate the concentration of NV^- centres to be $1.1 \times 10^{18} \text{ cm}^{-3}$.

$\sim 12 \text{ mK}$ while keeping the qubit at its optimum flux bias (degeneracy point). The total Hamiltonian of the coupled system is

$$H = \frac{h}{2}(\Delta\sigma_x + \varepsilon\sigma_z) + h \sum_i \left[DS_{z,i}^2 + E(S_{x,i}^2 - S_{y,i}^2) \right] + \frac{h}{2} \sum_i g_i \sigma_z S_{x,i}$$

where i runs over the NV^- centres that couple to the flux qubit. The corresponding coupling constant can be estimated using the Biot-Savart law as $g_i \approx 8.8 \text{ kHz}$ (see Supplementary Information, sections III and IV, for further details). In this set-up, states $|m_s = \pm 1\rangle_i$ of the NV^- electronic spin are nearly degenerate and our flux qubit therefore couples to both the $|0\rangle_i \rightleftharpoons |1\rangle_i$ transition and the $|0\rangle_i \rightleftharpoons |-1\rangle_i$ transition. This results in an effective coupling constant of $\sqrt{2}g_i$ when a degenerate NV^- centre is treated as a two-level system.

From the spectroscopic measurements, a clear anticrossing was observed (Fig. 3a) near the degeneracy point of the flux qubit, whereas no such gap was observed in the same flux qubit before the mounting of the ensemble (Fig. 3a, inset). We also note a narrow resonance, of width less than 1 MHz , at 2.878 GHz near this anticrossing. This can be ascribed to the near-degenerate excited states of the NV^- ensemble. From the high-resolution spectrum shown in Fig. 3b, a vacuum Rabi

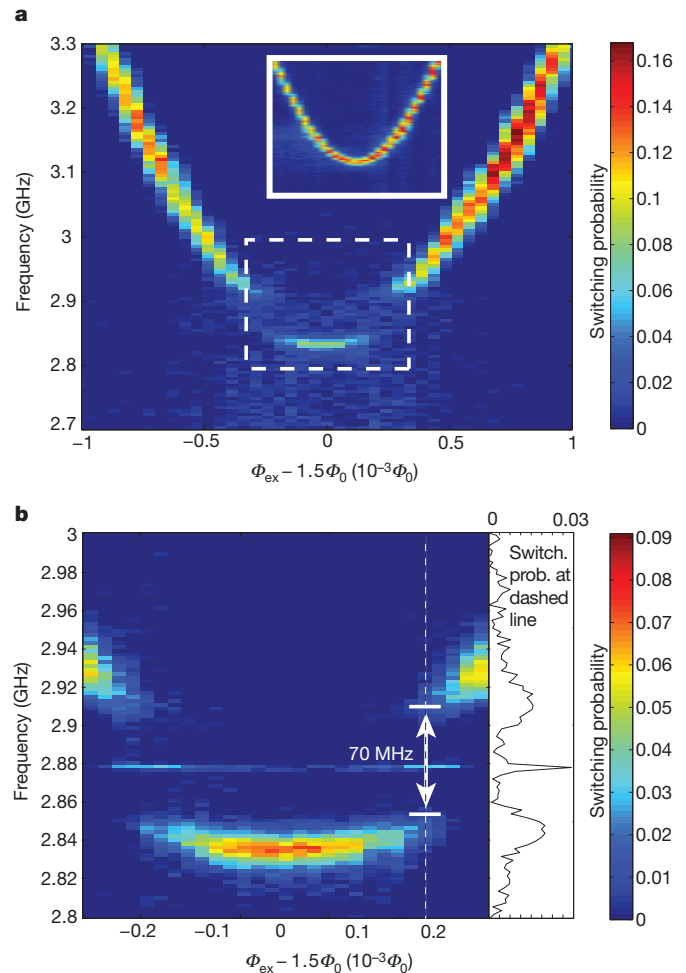


Figure 3 | Energy spectrum of the flux qubit coupled to a NV^- ensemble. **a**, Resonant frequencies indicated by the SQUID detector switching probability (when a 500-ns microwave pulse excites the system before the read-out pulse) versus the external magnetic flux, $\Phi_{\text{ex}} = \Phi_m + \Phi_\alpha/2$, where Φ_m and Φ_α are the fluxes through the qubit main loop and the qubit α -loop, respectively. Inset, spectrum over the same region before mounting of the diamond crystal. **b**, Magnified view of the dashed region in **a**, including a cross-sectional view of the cut along the dashed line. A vacuum Rabi splitting of 70 MHz is observed. Because the qubit phase relaxation time is $T_2^{\text{cho}} \approx 0.25 \mu\text{s}$, we are operating in the strong-coupling regime.

splitting near $g_{\text{ens}} \approx 70$ MHz was clearly observed, confirming strong coupling (two broad resonance peaks are present in the cross-sectional view) between the flux qubit and the NV^- ensemble. The ensemble can be seen as an effective harmonic oscillator strongly coupled to the flux qubit¹⁷. Next, from the measured vacuum Rabi splitting and our calculated value of g_{e} , we can estimate the number of NV^- centres in the coupled ensemble to be $N = g_{\text{ens}}^2 / 2g^2 \approx 3.2 \times 10^7$, where the factor of two in the denominator is due to the twofold degeneracy of the excited states $|\pm 1\rangle_i$ of a NV^- centre. This estimate is consistent with the density of NV^- centres measured by photoluminescence spectroscopy of the whole sample ($1.1 \times 10^{18} \text{ cm}^{-3}$) multiplied by the volume of centres coupling to the flux qubit ($40 \mu\text{m}^2$ (area) $\times 0.7 \mu\text{m}$ (effective thickness)). This spectroscopy approach gives the number of coupled centres as $\sim 3.1 \times 10^7$.

Next we investigated the dynamics of our system in the time domain, with a measurement cycle similar to that used in systems of qubits coupled with LC (inductor–capacitor) resonators²⁵ (Fig. 4a, inset). We first excited the flux qubit and then brought it into resonance with the NV^- ensemble. Single-energy quantum exchange between the flux qubit and the NV^- ensemble at resonance manifests itself as the vacuum Rabi oscillations

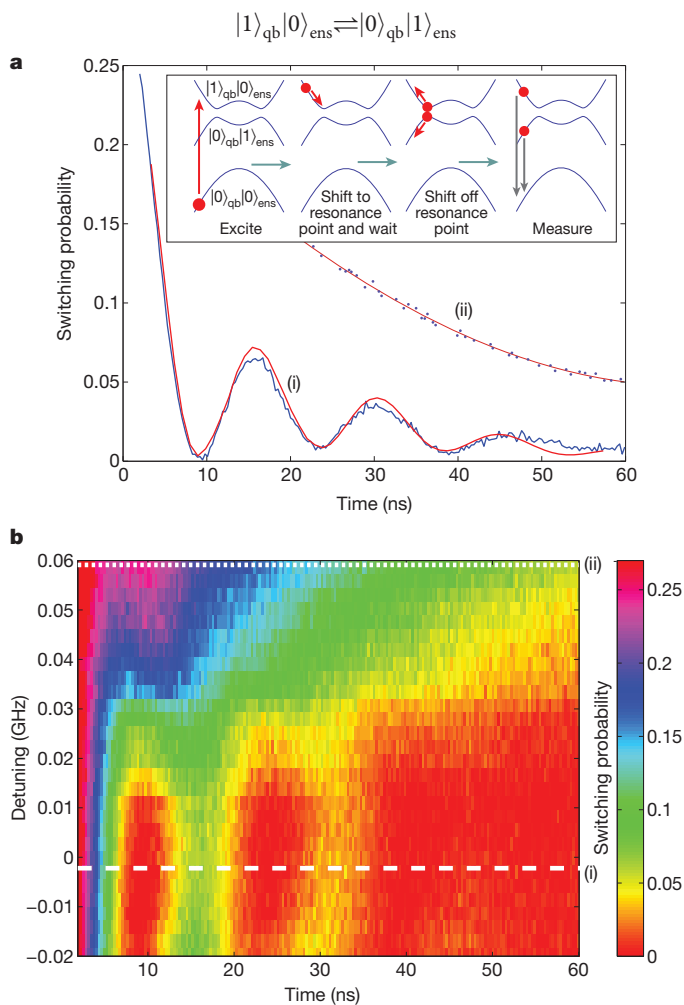


Figure 4 | Vacuum Rabi oscillations of the flux qubit/ NV^- ensemble coupled system. **a**, Line (i): damped oscillation (blue) of an initially excited flux qubit resonantly coupled to a NV^- ensemble. The red line shows the results of a numerical model (Supplementary Information, section III). Line (ii): the same sequence as in line (i), but with 60-MHz detuning from resonance. Inset, measurement sequence. **b**, Two-dimensional plot of the SQUID detector switching probability as a function of both time and the detuning by flux bias shift. The white dashed and dotted lines correspond to the switching probabilities shown in **a** (as indicated).

where $|0\rangle_{\text{qb}}$ and $|1\rangle_{\text{qb}}$ are respectively the ground and excited states of the qubit; $|0\rangle_{\text{ens}} = |00 \dots 0\rangle$ is the ground state of the ensemble, with each individual NV^- centre in the state $|m_S = 0\rangle$; and $|1\rangle_{\text{ens}}$ is a Dicke state of the spin ensemble with one excitation, written as $|1\rangle_{\text{ens}} = (1/\sqrt{N}) \sum_i S_{+,i} |00 \dots 0\rangle$. Here the operator $S_{+,i} = (1/\sqrt{2})[|1\rangle_i \langle 0|_i + |-1\rangle_i \langle 0|_i]$ excites the i th NV^- spin into a superposition of states $|1\rangle$ and $|-1\rangle$.

Figure 4a shows vacuum Rabi oscillations between the flux qubit and the ensemble of electronic spins at the 2.878-GHz resonance. However, the decay time of the oscillations is approximately 20 ns. This is much shorter than the relaxation times of both the flux qubit ($T_{1,\text{qb}} \approx 150$ ns) and the NV^- ensemble ($T_{1,\text{NV}} \gg 10 \mu\text{s}$). We cannot improve the vacuum Rabi decay time even by tuning the qubit degeneracy point closer to the NV^- resonance frequency. As we tune the flux qubit away from the 2.878-GHz resonance, by changing Φ_{m} (Supplementary Information, section I), the coherence time associated with this measurement scheme becomes considerably longer (Fig. 4a, line (ii), and Fig. 4b). From these results, we conclude that a strong source of dephasing of unknown origin exists in the system near resonance.

There are several likely sources of dephasing. The most probable is a large electron spin-half bath derived from the P1 centres (a nitrogen atom substituting a carbon atom) present in the HPHT-synthesized type-Ib diamond crystal used to prepare the NV^- ensemble. In our situation, where there is no external magnetic field, the NV^- centres and P1 centres naturally couple¹², in which case an enhanced decay may result²⁶. The effect of P1 centres can be eliminated to a large extent by applying an external magnetic field to split the NV^- states $|\pm 1\rangle$. A 1-mT field could split these by approximately 15 MHz, detuning them relative to the P1 centres and thus considerably improving the dephasing time of the coupled system. We can also decrease the number of P1 centres in the sample (from 100 p.p.m. to 1 p.p.m.) by using differently synthesized diamond crystals. In addition, by using non-HPHT-synthesized type-Ib crystals we can remove the effect of other natural defects that may be present. Finally, there is also a strong hyperfine interaction (~ 100 MHz) between the NV^- electron spin and ^{13}C nuclear spins. If the nuclear spins are not initially polarized, unwanted dephasing will result. Polarizing the nuclear spins, thus removing this source of dephasing, should allow us to observe vacuum Rabi oscillations where we are limited by the dephasing time of the flux qubit.

In conclusion, we have experimentally demonstrated strong coherent coupling between a flux qubit and an ensemble of nitrogen-vacancy colour centres in single-crystal diamond. Furthermore, we have observed, via vacuum Rabi oscillations, the coherent exchange (transfer) of a single quantum of energy. This is the first step towards the realization of a long-lived quantum memory for condensed-matter systems, with an additional potential future application as an interface between the microwave and optical domains.

METHODS SUMMARY

Photoluminescence of the NV^- centres. A confocal microscope was used to measure the photoluminescence of NV^- centres with excitation from a Nd:YAG laser²⁷ (wavelength, 532 nm). The photon antibunching was measured using a Hanbury Brown/Twiss set-up to identify the single NV^- centre. In the confocal microscopy, the signal intensity depends on the number of colour centres in the detected spot region, which is limited by the wavelength of the incident light and the numerical aperture of the objective lens. From the ratio, P , of the respective photoluminescence intensities of the single NV^- centre and the ensemble, the concentration of the ensemble of the NV^- centres, C , can be estimated from $C = P/V$, where V is the volume of the detected region ($5.85 \times 10^7 \text{ nm}^3$). We measured the ratio to be $P = (6.45 \pm 0.30) \times 10^4$, and so estimated C to be $1.1 \times 10^{18} \text{ cm}^{-3}$.

Spectroscopy of the flux qubit. To make spectroscopic measurements, we used a 500-ns-long microwave pulse to excite the flux qubit. This was followed by a read-out pulse on the d.c. SQUID that determines the detector's switching probability and allows us to infer the state of the flux qubit.

Received 30 May; accepted 10 August 2011.

1. Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).
2. Nakamura, Y., Pashkin, YuA & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
3. Vion, D. *et al.* Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
4. Chiorescu, I., Nakamura, Y., Harmans, C. J. P. M. & Mooij, J. E. Coherent quantum dynamics of a superconducting flux qubit. *Science* **299**, 1869–1871 (2003).
5. Sillanpää, M. A., Park, J. I. & Simmonds, R. W. Coherent quantum state storage and transfer between two phase qubits via a resonant cavity. *Nature* **449**, 438–442 (2007).
6. Majer, J. *et al.* Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
7. DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
8. Ansmann, M. *et al.* Violation of Bell's inequality in Josephson phase qubits. *Nature* **461**, 504–506 (2009).
9. Sørensen, A. S. van der Wal, C. H., Childress, L. I. & Lukin, M. D. Capacitive coupling of atomic systems to mesoscopic conductors. *Phys. Rev. Lett.* **92**, 063601 (2004).
10. Tian, L., Rabl, P., Blatt, R. & Zoller, P. Interfacing quantum-optical and solid-state qubits. *Phys. Rev. Lett.* **92**, 247902 (2004).
11. Rabl, P. *et al.* Hybrid quantum processors: molecular ensembles as quantum memory for solid state circuits. *Phys. Rev. Lett.* **97**, 033003 (2006).
12. Marcos, D. *et al.* Coupling nitrogen-vacancy centers in diamond to superconducting flux qubits. *Phys. Rev. Lett.* **105**, 210501 (2010).
13. Brune, M. *et al.* Quantum Rabi oscillation: a direct test of field quantization in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).
14. Chiorescu, I., Groll, N., Bertaina, S., Mori, T. & Miyashita, S. Magnetic strong coupling in a spin-photon system and transition to classical regime. *Phys. Rev. B* **82**, 024413 (2010).
15. Wu, H. *et al.* Storage of multiple coherent microwave excitations in an electron spin ensemble. *Phys. Rev. Lett.* **105**, 140503 (2010).
16. Schuster, D. I. *et al.* High-cooperativity coupling of electron-spin ensembles to superconducting cavities. *Phys. Rev. Lett.* **105**, 140501 (2010).
17. Kubo, Y. *et al.* Strong coupling of a spin ensemble to a superconducting resonator. *Phys. Rev. Lett.* **105**, 140502 (2010).
18. Amsüss, R. *et al.* Cavity QED with magnetically coupled collective spin states. *Phys. Rev. Lett.* **107**, 060502 (2011).
19. Raizen, M. G., Thompson, R. J., Brecha, R. J., Kimble, H. J. & Carmichael, H. J. Normal-mode splitting and linewidth averaging for two-state atom in an optical cavity. *Phys. Rev. Lett.* **63**, 240–243 (1989).
20. Naydenov, B. *et al.* Enhanced generation of single optically active spins in diamond by ion implantation. *Appl. Phys. Lett.* **96**, 163108 (2010).
21. Neumann, P. *et al.* Excited-state spectroscopy of single NV defects in diamond using optically detected magnetic resonance. *N. J. Phys.* **11**, 013017 (2009).
22. Zhu, X., Kemp, A., Saito, S. & Semba, K. Coherent operation of a gap-tunable flux qubit. *Appl. Phys. Lett.* **97**, 102503 (2010).
23. Fedorov, A. *et al.* Strong coupling of a quantum oscillator to a flux qubit at its symmetry point. *Phys. Rev. Lett.* **105**, 060503 (2010).
24. Mooij, J. E. *et al.* Josephson persistent-current qubit. *Science* **285**, 1036–1039 (1999).
25. Johansson, J. *et al.* Vacuum Rabi oscillations in a macroscopic superconducting qubit LC oscillator system. *Phys. Rev. Lett.* **96**, 127006 (2006).
26. Hanson, R., Dobrovitski, V. V., Feiguin, A. E., Gywat, O. & Awschalom, D. D. Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science* **320**, 352–355 (2008).
27. Mizuochi, N. *et al.* Coherence of single spins coupled to a nuclear spin bath of varying density. *Phys. Rev. B* **80**, 041201(R) (2009).
28. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. Observation of coherent oscillations in a single electron spin. *Phys. Rev. Lett.* **92**, 076401 (2004).
29. Gruber, A. *et al.* Scanning confocal optical microscopy magnetic resonance on single defect centers. *Science* **276**, 2012–2014 (1997).
30. Kurtsiefer, Ch, Zarda, P., Mayer, S. & Weinfurter, H. A stable solid-state source of single photons. *Phys. Rev. Lett.* **85**, 290–293 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Tawara, H. Gotoh and T. Sogawa for optical measurements at an early stage of this work. We also thank H. Tanji, Y. Matsuzaki, S. J. Devitt, J. Schmiedmayer and J. E. Mooij for discussions. This work was supported in part by the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST), Scientific Research of Specially Promoted Research (grant no. 18001002) by MEXT, a Grant-in-Aid for Scientific Research on Innovative Areas (grant no. 22102502), and Scientific Research (A) grant no. 22241025 from the Japanese Society for the Promotion of Science (JSPS). M.S.E. was supported by a JSPS fellowship.

Author Contributions All authors contributed extensively to the work presented in this paper. X.Z. and A.K. carried out measurements and data analysis on the coupled flux qubit/NV⁻ ensemble. N.M., M.K. and K.S. prepared and characterized the NV⁻ diamond crystals. X.Z., S.K., S.S. and A.K. designed and fabricated the flux qubit and associated devices. S.S., K.K. and A.K. designed and developed the flux qubit measurement system. W.J.M., A.K., Y.T., H.N., M.S.E. and K.N. provided theoretical support and analysis. X.Z., M.S.E., W.J.M. and K.S. wrote the manuscript, with feedback from all authors. W.J.M. and K.S. designed and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.S. (semba.kouichi@lab.ntt.co.jp).

Self-replication of information-bearing nanoscale patterns

Tong Wang¹, Ruojie Sha¹, Rémi Dreyfus^{2†}, Mirjam E. Leunissen^{2†}, Corinna Maass², David J. Pine², Paul M. Chaikin² & Nadrian C. Seeman¹

DNA molecules provide what is probably the most iconic example of self-replication—the ability of a system to replicate, or make copies of, itself. In living cells the process is mediated by enzymes and occurs autonomously, with the number of replicas increasing exponentially over time without the need for external manipulation. Self-replication has also been implemented with synthetic systems, including RNA enzymes designed to undergo self-sustained exponential amplification^{1–5}. An exciting next step would be to use self-replication in materials fabrication, which requires robust and general systems capable of copying and amplifying functional materials or structures. Here we report a first development in this direction, using DNA tile motifs that can recognize and bind complementary tiles in a pre-programmed fashion. We first design tile motifs so they form a seven-tile seed sequence; then use the seeds to instruct the formation of a first generation of complementary seven-tile daughter sequences; and finally use the daughters to instruct the formation of seven-tile granddaughter sequences that are identical to the initial seed sequences. Considering that DNA is a functional material that can organize itself and other molecules into useful structures^{6–13}, our findings raise the tantalizing prospect that we may one day be able to realize self-replicating materials with various patterns or useful functions.

Nucleic acids comprise the genetic material of all living organisms, with the ordered pattern of bases in the DNA double helix readily copied by an enzymatic process that leads to semi-conservative replication inside the cell^{14,15}. The replication process exploits the complementarity of the four-letter code of bases in the linear DNA double helix, where adenine (A) pairs with thymine (T) and guanine (G) pairs with cytosine (C)¹⁶. We transfer this concept to DNA materials, and use as an analogue of the nucleotide letter the bent triple crossover motif (BTX)¹⁷, which displays four single strands (of seven nucleotides each) to connect with a second, similar BTX molecule. Information is encoded using these four strands, which ensure a specific interaction between two BTX molecules and therefore the transfer of information from a seed structure to later generations. As shown in Fig. 1a, the structure formed by two paired BTX letters is a paired 6-helix bundle motif (P6HB). It contains six helices connected to each other by four small double helices that extend from one duplex to another.

Once specific recognition has been set and BTX motifs have been paired on an initial seed as desired, they must be bound together linearly, analogous to the formation of a DNA backbone. This is achieved using 18 nucleotides per duplex (9 from each side of the three BTX double-helical domains) and binding them with a solution-derived DNA strand. We use the same DNA sequence in every case (except in the seed), so that any pattern can be replicated. The cohesive strength of the longitudinal bonds linking our tiles to form longer arrays is necessarily of higher melting temperature than the lateral binding between seed and daughter arrays, both because of greater length and because of end-stacking interactions¹⁸; this is analogous to the bases in the double helix being strongly attached (covalently

bonded) in the direction parallel to the helix axis and weakly attached (hydrogen-bonded) perpendicular to it. The difference in bonding interactions allows a complementary sequence to be assembled on a seed template, internally fused, and then separated from the template to form an autonomous copy. We note that, in contrast to the DNA code, the BTX-based code is not limited to four letters: in principle, we can design as many as 4²⁸ different combinations of strands involved in lateral pairing.

Experimentally, we start with the simplest case of using two different logical BTX species: an A tile, a B tile, and their complementary A' and B' tiles. Tiles are labelled for identification by atomic force microscopy (AFM), with either biotinylated nucleotides that bind streptavidin or

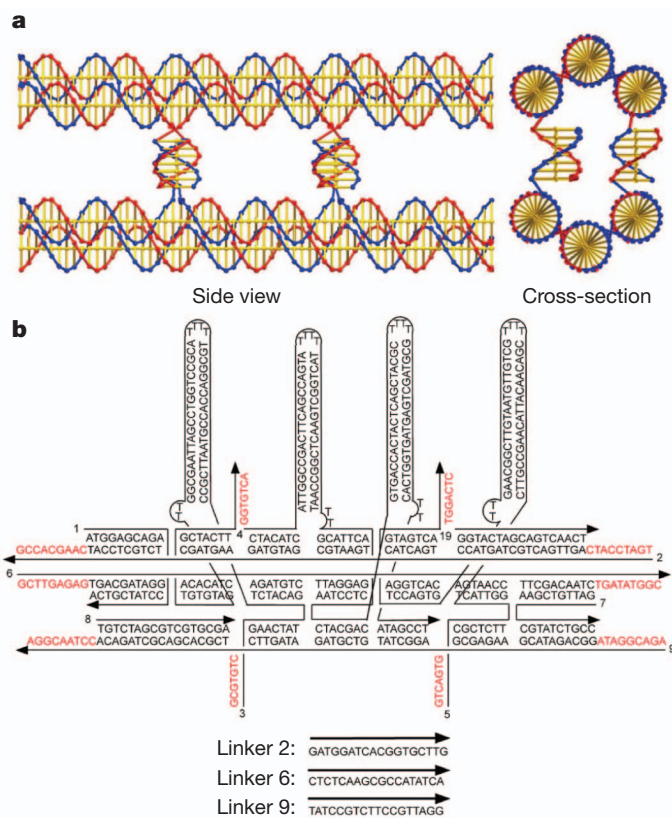


Figure 1 | DNA tile sequences and structures. **a**, The P6HB motif, drawn with GIDEON²⁴, comprising two BTX domains paired by four lateral connections. The cross-section view shows two of the four helices that are formed by the lateral cohesive interactions. The connections at the rear are eclipsed in this projection. **b**, Sequence and structure of the B' BTX tile. Four helical domains, hairpins, are attached perpendicular to the BTX motif, to create a topographic feature that can be detected by AFM. Four lateral and six longitudinal sticky ends are shown in red. Other tiles are shown in Supplementary Information section 1.

¹Department of Chemistry, New York University, New York, New York 10003, USA. ²Center for Soft Matter Research, New York University, New York, New York 10003, USA. [†]Present addresses: Complex Assemblies of Soft Matter, CNRS-Rhodias-UPenn UMI 3254, Bristol, Pennsylvania 19007, USA (R.D.); FOM Institute AMOLF, Sciencepark 104, 1098 XG Amsterdam, The Netherlands (M.E.L.).

with large hairpin features (see Fig. 1b for a labelled B' tile). We can thus read with AFM any sequence created from A and B components, and check the fidelity of our replication from the initial seed to the new generations of daughters and granddaughters. The labelling also allows us to verify whether the initial seed pattern has formed accurately. The first tile of the seed is an initiator I (an A-like tile attached to a magnetic bead), followed by a series of A and B tiles to give the sequence IBBABAB. In the seed it is the A tiles that contain labels, whereas in the complements it is the B' tiles. Thus, using L and N to represent, respectively, the presence and absence of labels, the seed array should have the sequence LNNLNLN and the newly replicated sequence should appear as NLLNLNL. (Note, however, that any array can be labelled uniquely in any generation.) In addition to creating the seed and its replica, we have also replicated the replica and thereby obtained a direct copy of the information in the seed array itself.

The starting point for each stage in the replication process is the self-assembly of BTX tiles. (See Supplementary Information section 1 for the sequences of every tile used; section 2 for non-denaturing gels that demonstrate that the BTX tiles form properly and only associate as complementary pairs; and section 3 for electrophoretic mobility (Ferguson) analyses of the individual tiles and their P6HB complexes.)

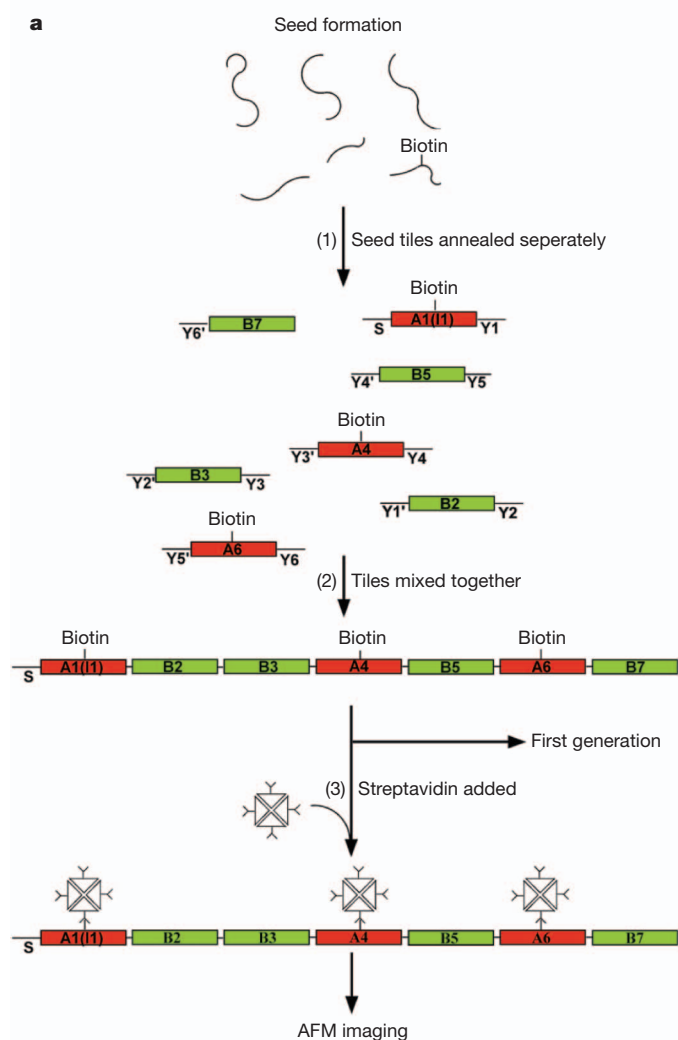
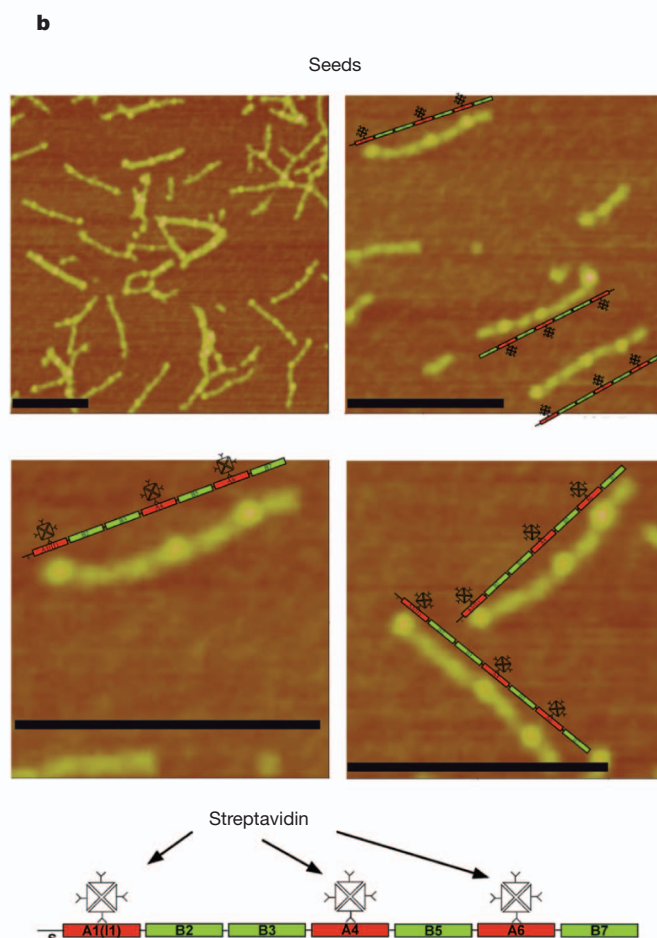


Figure 2 | DNA seeds. **a**, Seed formation. In step 1, the individual strands of the seed tiles self-assemble in separate vessels to produce seven different BTX tiles, each flanked by unique sticky ends labelled Y and a number; primed numbers are complementary to unprimed numbers. The red tiles are the A tiles and the green tiles are the B tiles. The A tiles contain a biotin group to enable their decoration by streptavidin. The tile labelled A1(I1) is the initiator tile. The strand labelled S on its left can bind to a dynabead during the replication

process. In step 2, the tiles are mixed together, producing 7-unit seeds when they are mixed together. In step 3, the tiles are prepared for AFM imaging by the addition of streptavidin. **b**, AFM images of seeds. The upper left panel shows a typical field of view, of slightly less than $1 \mu\text{m}^2$. A large number of seeds are present, along with some multimeric complexes. The other three panels are zoomed images. (Black scale bars, 200 nm.) A schematic image of each seed is shown next to the seed.



process. In step 2, the tiles are mixed together, producing 7-unit seeds when they are mixed together. In step 3, the tiles are prepared for AFM imaging by the addition of streptavidin. **b**, AFM images of seeds. The upper left panel shows a typical field of view, of slightly less than $1 \mu\text{m}^2$. A large number of seeds are present, along with some multimeric complexes. The other three panels are zoomed images. (Black scale bars, 200 nm.) A schematic image of each seed is shown next to the seed.

the self-assembly of A' and B' motifs, annealed in different vessels (step 1), and then added to a solution of seed arrays and mixed (step 2). The initiator tile of the daughter generation contains the same S end that was on the initiator tile of the seed, but this strand is protected by a cover strand¹⁹ to keep it free of any beads. Otherwise, the tiles are all flanked by the same set of sticky-end connectors (nonamers denoted Y and Z in the figure). The B' tiles contain a group of covalently attached hairpins as labels (Fig. 1b).

Following mixing, the BTX tiles with complementary lateral cohesive sequences are paired together to form an extended P6HB motif; the

result is a complementary daughter array attached to the seed array, with the daughter tiles not yet bound to each other. At this point, a dynabead is introduced, containing a linker complementary to the S strand on the seed, so that the entire construct is attached to the bead. Following a wash, 18-mer linkers (Fig. 1b), complementary to both of the nonamer sticky ends extending between tiles, are added to the solution. These 18-mers are annealed to the construct, connecting the tiles (magnified inset following step 7 in Fig. 3a). Following a wash to remove excess linkers, seeds and daughters are separated by heating the system to 37 °C, at which temperature lateral pairs

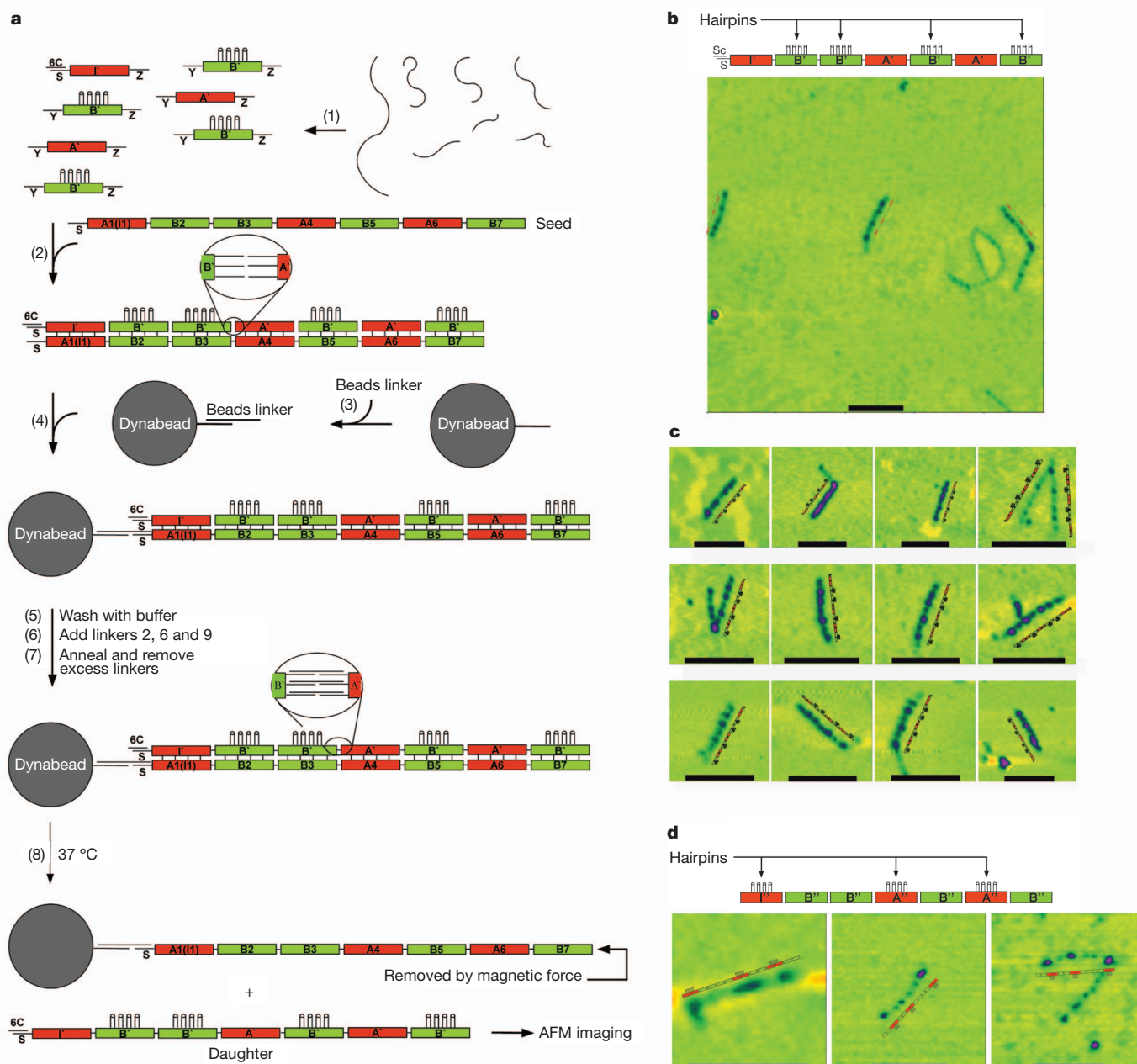


Figure 3 | DNA generations. **a**, Replication of the seven-tile seed pattern in the first generation. In step 1, strands are annealed into tiles, all flanked by the same connectors, designated Y and Z. The initiator tile (I') contains a protected S-strand, paired with a cover strand, 6C. The B' tiles (Fig. 1b) contain the 4-hairpin markers for AFM imaging. In the presence of the seed (step 2), the strands assemble into a pattern mimicking the seed pattern. The magnetic dynabead (large grey circle) is prepared in step 3, and attached to the seed (step 4). This is followed by a wash step, the addition of linkers, and their annealing

(steps 5–7). Heating the system to 37 °C results in the separation of the daughter 7-tile complex and the seed (removed magnetically, using the dynabead). **b**, **c**, AFM images of first-generation daughter complexes, showing a typical field of view (**b**) and a collection of zoomed images (**c**). (Black scale bars, 200 nm.) These images, adjacent to explanatory schematics, demonstrate that the ABBABAB pattern has been replicated successfully. **d**, AFM images of second-generation molecules, showing the pattern that was programmed in the original seed tile.

are broken but the longitudinal bonds remain intact (step 8, bottom of Fig. 3a).

Despite the large number of manipulations required to prepare the daughters, the large-field AFM image of this system (Fig. 3b) still contains multiple copies of the molecules in an area measuring a little over $1\ \mu\text{m}^2$, along with some flawed products. Comparison with the image in Fig. 2b reveals that the daughters are significantly sparser than the seeds. The zoomed images in Fig. 3c show daughter tile arrays with four clear features (the hairpin labels), showing that the arrays appear to have formed correctly, although some of them are parts of aggregates. The percentage of heptamers with the correct sequence is estimated to be 70% (23 molecules sampled).

Granddaughters (replicas of the seed arrays) were produced directly from an initial preparation, rather than from purified seeds or purified daughters (see Supplementary Information section 4). Cover strands were removed from the initiator tile, and A' and B' tiles were added to the solution. In this second generation, A' tiles contain the hairpin labels. The same sequence of steps is then performed as in the generation of the daughter strands (see Methods and Supplementary Information, section 4), with daughters and granddaughters separated by heating to 37°C .

Figure 3d shows AFM images of granddaughters, of which there are fewer molecules and fewer cases of unambiguously correct images than seen with daughters. We estimate the percentage of second-generation heptamers with a correct tile sequence to be 31% (18 molecules sampled), corresponding to 55% ($\sim 0.31^{1/2}$) per generation, similar to the 70% seen for the daughters (See Supplementary Information section 5). We note that a net increase of molecules, and hence a growing system, would be obtained if we were to retain the original seeds in the solution containing the granddaughter molecules.

At present, our replication method is rather cumbersome, as it requires multiple chemical and thermal processing cycles; and, unlike the work of ref. 1 on RNA enzymes¹, it does not yet achieve exponential amplification. However, we have demonstrated that it is possible to replicate not just molecules such as DNA or RNA, but discrete tertiary structures that could in principle assume many different shapes and functional features. We also note that, in the same way as the first cars, airplanes and computers were clunky relative to current systems, it should be possible to make the present procedure smoother and more sophisticated. For example, yield is obviously affected by the removal of seed molecules to produce the daughters, and likewise by the removal of daughters to produce the granddaughters so as to simplify analysis. This could be avoided through elimination of bead removal steps, by using self-protected hairpins²⁰ or photoactive molecules for the longitudinal interactions. We expect that this and other improvements will deliver a robust replication method that is applicable to molecular, nanometre-sized and colloidal systems (such as patchy²¹ and lock and key²² particles) displaying programmed recognition.

METHODS SUMMARY

All strands were designed using the program SEQUIN²³. Following PAGE purification, strands for the seeds, daughter and granddaughter tiles were mixed stoichiometrically as estimated by OD₂₆₀ and dissolved to $0.5\ \mu\text{M}$ in TAE/Mg²⁺ buffer (40 mM Tris-HCl, 20 mM acetic acid, 2 mM EDTA, 12.5 mM magnesium acetate, pH 8.0). The solutions were slowly annealed from 90°C to 23°C over 48 h in a 2-litre water bath insulated in a Styrofoam box. Stoichiometric quantities of seven seed tiles were mixed and annealed from 45°C to 23°C over 24 h to make seeds. To form the first generation, three first-generation tiles (I', A' and B') were mixed with annealed first-generation tiles (seeds: I':A':B' = 1:2:4:8), and slowly annealed from 45°C to 23°C . Dynabeads were washed with double-distilled H₂O and TAE/Mg buffer, mixed with beads linker in TAE/Mg buffer, slowly annealed from 55°C to 23°C , washed with buffer, and mixed with DNA solution. The solution containing dynabeads was annealed from 33°C to 23°C , placed on a magnetic stand and washed with TAE/Mg buffer. Linking strands 2, 6 and 9 (Fig. 1b) were then added, the solution cooled from 33°C to 23°C , placed on a magnetic stand and washed with TAE/Mg buffer to remove excess linkers. Dynabeads in TAE/Mg buffer were kept at 37°C for one hour, placed on the magnetic stand, and the solution was removed from dynabeads and stored in a clean tube for AFM imaging. Formation of the second generation is similar

to the first: it starts from initial seed preparation, followed by formation of the first generation, and adding second-generation tiles (I'', A'' and B''). Steps 2–8 described in formation of the first generation were repeated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 October 2010; accepted 18 August 2011.

- Lincoln, T. A. & Joyce, G. F. Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232 (2009).
- Wintner, E. A., Conn, M. M. & Rebek, J. Jr. Studies in molecular replication. *Acc. Chem. Res.* **27**, 198–203 (1994).
- Schulman, R. & Winfree, E. Synthesis of crystals with a programmable kinetic barrier to nucleation. *Proc. Natl Acad. Sci. USA* **104**, 15236–15241 (2007).
- Lin, C. *et al.* In vivo cloning of artificial DNA nanostructures. *Proc. Natl Acad. Sci. USA* **105**, 17626–17631 (2008).
- Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of symbiosis in peptide self-replication through a peptide hypercyclic network. *Nature* **390**, 591–594 (1997).
- Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
- Zhu, L., Lukeman, P. S., Canary, J. W. & Seeman, N. C. Nylon/DNA: single-stranded DNA with covalently stitched nylon lining. *J. Am. Chem. Soc.* **125**, 10178–10179 (2003).
- Zheng, J. *et al.* 2D nanoparticle arrays show the organizational power of robust DNA motifs. *Nano Lett.* **6**, 1502–1504 (2006).
- Nykypanchuk, D., Maye, M. M., van der Lelie, D. & Gang, O. DNA-guided crystallization of colloidal nanoparticles. *Nature* **451**, 549–552 (2008).
- Park, S. Y. *et al.* DNA-programmable nanoparticle crystallization. *Nature* **451**, 553–556 (2008).
- Kim, A. J., Biancaniello, P. L. & Crocker, J. C. Engineering DNA-mediated colloidal crystallization. *Langmuir* **22**, 1991–2001 (2006).
- Valignat, M. P., Theodoly, O., Crocker, J. C., Russel, W. B. & Chaikin, P. M. Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids. *Proc. Natl Acad. Sci. USA* **102**, 4225–4229 (2005).
- Leunissen, M. E. *et al.* Towards self-replicating materials of DNA-functionalized colloids. *Soft Matter* **5**, 2422–2430 (2009).
- Watson, J. D. & Crick, F. H. C. The genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967 (1953).
- Lehman, I. R., Bessman, M. J., Simms, E. S. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acids 1. *J. Biol. Chem.* **233**, 163–170 (1958).
- Watson, J. D. & Crick, F. H. C. The molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
- Kuzuya, A., Wang, R., Sha, R. & Seeman, N. C. Six-helix and eight-helix DNA nanotubes assembled from half-tubes. *Nano Lett.* **7**, 1757–1763 (2007).
- Wang, R., Kuzuya, A., Liu, W. & Seeman, N. C. Blunt-ended DNA stacking interactions in a 3-helix motif. *Chem. Commun.* **46**, 4905–4907 (2010).
- Zhong, H. & Seeman, N. C. RNA used to control a DNA rotary machine. *Nano Lett.* **6**, 2899–2903 (2006).
- Leunissen, M. E. *et al.* Switchable self-protected attractions in DNA-functionalized colloids. *Nature Mater.* **8**, 590–595 (2009).
- Zhang, Z. & Glotzer, S. C. Self-assembly of patchy particles. *Nano Lett.* **4**, 1407–1413 (2004).
- Sacanna, S., Irvine, W. T. M., Chaikin, P. M. & Pine, D. J. Lock and key colloids. *Nature* **464**, 575–578 (2010).
- Seeman, N. C. *De novo* design of sequences for nucleic acid structure engineering. *J. Biomol. Struct. Dyn.* **8**, 573–581 (1990).
- Birac, J. J., Sherman, W. B., Kopatsch, J., Constantinou, P. E. & Seeman, N. C. GIDEON. A program for design in structural DNA nanotechnology. *J. Mol. Graph. Model.* **25**, 470–480 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was partially supported by the W.M. Keck Foundation (P.M.C., D.J.P., N.C.S.), the National Science Foundation (MRSEC Program award DMR-0820341, P.M.C., D.J.P. and grants CTS-0608889 and CCF-0726378, to N.C.S.), NASA (NNX08AK04G, to P.M.C.), the National Institute of General Medical Sciences (grant GM-29554, to N.C.S.), the Army Research Office (grants 48681-EL and W911NF-07-1-0439, to N.C.S.), the Office of Naval Research (grants N000140910181 and N000140911118, to N.C.S.), the Netherlands Organization for Scientific Research (Rubicon grant, to M.E.L.) and DAAD (post-doctoral grant, to C.M.).

Author Contributions T.W., R.S. and C.M. designed experiments, performed experiments, analysed data and wrote the paper; R.D. analysed data and wrote the paper; M.E.L. initiated the project, analysed data and wrote the paper; D.J.P. and P.M.C. initiated and directed the project, analysed data and wrote the paper; N.C.S. initiated and directed the project, designed experiments, analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.M.C. (chaikin@nyu.edu) or N.C.S. (ned.seeman@nyu.edu).

METHODS

DNA strand design, synthesis, and purification. DNA sequences were designed using the program SEQUIN²³. The DNA strands with modifications—for example, biotinylated—were synthesized on an Applied Biosystems 394, removed from the support, and deprotected using routine phosphoramidite procedures. Other strands were purchased from Integrated DNA Technologies, Inc. All strands were purified by denaturing gel electrophoresis (PAGE); bands were cut out of 15–20% denaturing gels and eluted in a solution containing 500 mM ammonium acetate, 11 mM magnesium acetate and 1 mM EDTA. This is the only purification step in the procedure.

Formation of hydrogen-bonded BTX complexes. The strands of each tile were mixed stoichiometrically, as estimated by OD₂₆₀, and dissolved to 0.5 μ M in TAE/Mg²⁺ buffer (40 mM Tris-HCl, 20 mM acetic acid, 2 mM EDTA, 12.5 mM magnesium acetate, pH 8.0). The solutions were slowly annealed from 90 °C to 23 °C over 48 h in a 2-l water bath insulated in a Styrofoam box.

Formation of self-replication seeds. Seven individual seed tiles were prepared using the protocol described above and were mixed stoichiometrically. The mixed solution was slowly annealed from 45 °C to room temperature over 24 h in a 2-l water bath insulated in a Styrofoam box. Seeds (seven tiles in length, in specific sequence) with biotin and without biotin were prepared separately in the same conditions. Seeds with biotin were used for AFM imaging only; seeds without biotin were used as starting material in the first step of self-replication. Before AFM imaging, seeds with biotin were incubated with streptavidin (biotin:streptavidin = 1:1) at room temperature for at least 6 h.

First step of self-replication (formation of first generation). (1) Three first-generation tiles (I', A' and B') were prepared as described above. (2) Annealed seeds were mixed with annealed first-generation tiles (seeds:I':A':B' = 1:2:4:8), slowly annealed from 45 °C to room temperature over 24 h in a 2-l water bath

insulated in a Styrofoam box. (3) Dynabeads (Invitrogen) were washed with ddH₂O and TAE/Mg buffer, mixed with beads linker in TAE/Mg buffer, slowly annealed from 55 °C to room temperature at 5 °C h⁻¹ on a rotator in a programmable incubator, and mixed with DNA solution prepared in step (2). (4) The above solution containing dynabeads was annealed from 33 °C to 23 °C at 1 °C h⁻¹ on a rotator in a programmable incubator. (5) Solution from (4) was placed on a magnetic stand for 2 min and washed with TAE/Mg buffer. (6) Linkers 2, 6 and 9 were added. (7) Solution was cooled from 33 °C to 23 °C at 1 °C h⁻¹ on a rotator in a programmable incubator, placed on a magnetic stand for 2 min, and washed with TAE/Mg buffer to eliminate excess linkers. (8) Dynabeads in TAE/Mg buffer were kept at 37 °C for 1 h, then placed on a magnetic stand for 2 min at 37 °C. The solution was removed from dynabeads and stored in a clean tube for AFM imaging.

Second step of self-replication (formation of second generation). Formation of the second generation starts from initial seed preparation, followed by formation of the first generation, but with steps (6) and (7), described above, performed before step (5). (1) Three second-generation tiles (I'', A'' and B'') were prepared as described above. A strand fully complementary to 6C was added to the solution of first-step self-replication and incubated at 25 °C for 2 h. (2) Steps (2)–(8) were performed as described for formation of first generation, except with steps (6) and (7) preceding step (5).

AFM imaging. All AFM imaging was performed in tapping-mode AFM in air. 5–7 μ l of DNA sample was spotted on freshly cleaved mica (Ted Pella, Inc.) and was left for 1 min to be absorbed. Mica was washed with 3–5 drops of ddH₂O three times, and excess water was removed by blotting the mica with filter paper. The mica was then blown dry using compressed air. All AFM imaging was performed on a NanoScope IV MultiMode SPM (Digital Instruments), using commercial cantilevers with silicon tips (VeecoProbes).

Atmospheric oxygenation caused by a change in volcanic degassing pressure

Fabrice Gaillard¹, Bruno Scaillet¹ & Nicholas T. Arndt²

The Precambrian history of our planet is marked by two major events: a pulse of continental crust formation at the end of the Archaean eon and a weak oxygenation of the atmosphere (the Great Oxidation Event) that followed, at 2.45 billion years ago. This oxygenation has been linked to the emergence of oxygenic cyanobacteria^{1,2} and to changes in the compositions of volcanic gases^{3,4}, but not to the composition of erupting lavas—geochemical constraints indicate that the oxidation state of basalts and their mantle sources has remained constant since 3.5 billion years ago^{5,6}. Here we propose that a decrease in the average pressure of volcanic degassing changed the oxidation state of sulphur in volcanic gases, initiating the modern biogeochemical sulphur cycle and triggering atmospheric oxygenation. Using thermodynamic calculations simulating gas–melt equilibria in erupting magmas, we suggest that mostly submarine Archaean volcanoes produced gases with $\text{SO}_2/\text{H}_2\text{S} < 1$ and low sulphur content. Emergence of the continents due to a global decrease in sea level and growth of the continental crust in the late Archaean then led to widespread subaerial volcanism, which in turn yielded gases much richer in sulphur and dominated by SO_2 . Dissolution of sulphur in sea water and the onset of sulphate reduction processes could then oxidize the atmosphere.

Mass-independent fractionation of sulphur (MIF), quantified as $\Delta^{33}\text{S}$, is commonly recorded in sediments older than 2.45 billion years (Gyr)⁷. A proliferation of MIF has been clearly identified between 2.7 and 2.45 Gyr, and older spikes are recorded at >3.4 Gyr (refs 7–9). MIF is now widely recognized as the signature of an anoxic atmosphere during the Archaean eon, which allowed the photolysis of atmospheric SO_2 conducive to MIF. The abrupt disappearance of MIF at 2.45 Gyr ago is thought to record the Great Oxidation Event—the rise of O_2 in the atmosphere that radically changed redox-based biogeochemical processes^{8–10}.

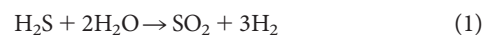
A more complete analysis¹¹ recently described how complex lithosphere–atmosphere–ocean transfers could have influenced the variation of Archaean MIF, particularly its strong rise between 2.7 and 2.45 Gyr ago, immediately preceding the Great Oxidation Event. A critical trigger of MIF could be the redox ratio of sulphur in volcanic gases ($\text{SO}_2/\text{H}_2\text{S}$) in an oxygen-poor atmosphere: high volcanic $\text{SO}_2/\text{H}_2\text{S}$ produces significant MIF, whereas a low ratio does not¹¹. It follows that a change to more-oxidized volcanic gases at the end of the Archaean could have caused the MIF explosion that preceded the Great Oxidation Event¹¹. The change in the redox state of volcanic gases has been attributed to evolution of the mantle source composition³, but the near invariance through geologic times of V/Sc and other redox-sensitive elements indicates that the redox state of the mantle and mantle-derived basalts has remained nearly constant from the middle Archaean to the present^{5,6}.

Kump and Barley¹² used Holland's redox criteria⁴ for volatile degassing to argue that subaerial volcanic gases are more oxidized than submarine ones and proposed that an increasing contribution from subaerial volcanism would have resulted in increasingly oxidizing volcanic emissions. But to define the redox state of Archaean

emissions, Kump and Barley¹² used fluids from seafloor hydrothermal vents whose compositions are controlled by redox-sensitive fluid–rock reactions¹³ and whose redox state, particularly during the Archaean, is poorly constrained^{4,14}. Furthermore, a survey of Hawaiian subaerial versus submarine volcanism¹⁵ indicates that submarine lavas are not more reduced than subaerial lavas. Our thermodynamic calculations show that the $\text{SO}_2/\text{H}_2\text{S}$ ratio of volcanic gases is only marginally influenced by redox conditions or pre-eruptive volatile contents of the lava, provided they remain within known limits. We show instead that degassing pressure controls this ratio.

Figure 1 reports the amount and speciation of magmatic sulphur in volcanic gases released from oceanic basalt calculated using a gas–melt equilibrium model¹⁶ (see Methods). For similar bulk volatile contents, variations of melt composition within the spectrum of terrestrial mafic magmas have only a minor effect on the composition of gas produced during decompression¹⁶. For simplicity we consider here a Hawaiian tholeiite whose degassing systematics have been well studied¹⁷. The composition of this basalt broadly mimics that of the tholeiites that dominate Archaean sequences. Figure 1 shows the effect of pressure and of the relative concentrations of pre-eruptive volatiles (S, H and C) on the amount and speciation of sulphur in volcanic gases. The black curves show degassing paths for an average basaltic source (oxygen fugacity, FMQ – 0.5 (0.5 log unit more reducing than the oxygen fugacity of the fayalite magnetite quartz (FMQ) buffer assemblage; 0.5 wt% H_2O ; 0.3 wt% CO_2 ; 0.15 wt% S). The total amount of sulphur in volcanic gases ($\text{H}_2\text{S} + \text{S}_2 + \text{SO}_2$) increases with decreasing pressure. At pressures exceeding 100 bar, the total degassed sulphur is relatively low, around 0.05 wt% for normally hydrated basalts. It reaches a maximum value at atmospheric pressure, where the degassed S species is in the range 0.17–0.25 wt%. The portion of degassed H_2S shows a maximum of 0.05–0.06 wt% at moderately low pressures in the range 20–100 bar. At still lower pressures, the quantity of H_2S dramatically decreases and SO_2 becomes the dominant species. At pressures higher than 20 bar, the average $\text{SO}_2/\text{H}_2\text{S}$ ratio is less than 0.5, reaching values of 2 to 8 at atmospheric pressure. Figure 1 therefore conveys two important factors, both independent of the source characteristics of the magma: (1) the total amount of degassed sulphur increases with decreasing pressure, and (2) at pressures exceeding a few tens of bar, sulphur is degassed as H_2S , whereas at atmospheric pressure, SO_2 is the dominant species.

Figure 2 shows how the overall composition of the gas phase varies as a function of pressure. SO_2 replaces H_2S as the dominant sulphur species at about 5 bar (Fig. 1 also shows that the pressure crossover ranges from 3 to 200 bar). The molar $\text{CO}_2/\text{H}_2\text{O}$ ratio also changes: deep degassing produces CO_2 -rich gases, whereas surface degassing produces water-rich gases. The H_2S – SO_2 shift is not directly related to the $\text{CO}_2/\text{H}_2\text{O}$ ratio but is essentially controlled by the following reaction, which occurs in the gas phase as pressure decreases (Methods):



Owing to the mole difference between reactants and products, a decrease in pressure displaces reaction (1) to the right. The shallow

¹Institut des Sciences de la Terre d'Orléans, CNRS-INSU/Université d'Orléans/Université de Tours, 1a rue de la Férolerie 45071, Orléans cedex 2, France. ²ISTerre, Université Joseph Fourier de Grenoble, CNRS, 1381 rue de la Piscine, 38400 Saint Martin d'Hères, France.

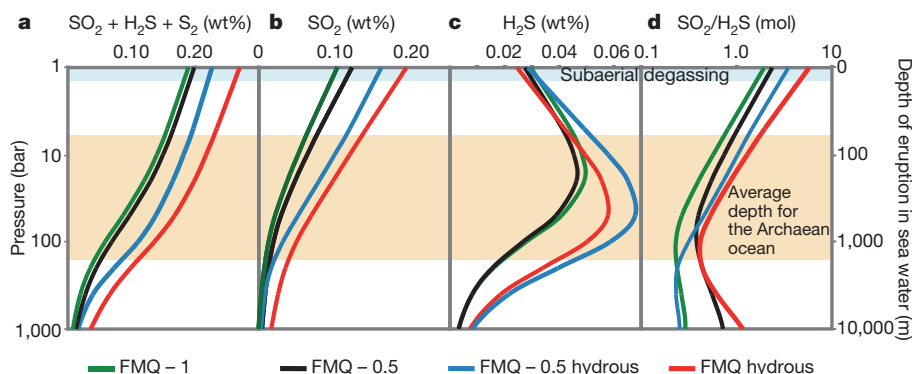


Figure 1 | Calculated sulphur content and speciation in volcanic gases as a function of pressure. We considered a basaltic magma containing $\text{H}_2\text{O} = 0.5$ wt%, $\text{CO}_2 = 0.3$ wt%, $\text{S} = 0.15$ wt% (black curves). Nevertheless, we varied the total water content ('hydrous' means 1 wt% H_2O) and the redox state (FMQ - 1 to FMQ), a range that encompasses conditions of subduction-related basalts and oceanic basalts, to illustrate that the influence of such

production of SO_2 and deep degassing of H_2S anticipated from theoretical considerations is confirmed by field volcanic gas studies^{16–18}. The production of SO_2 in the gas is accompanied by formation of H_2 (equation (1)), reaching 6 mol.% at atmospheric pressure. Figures 1 and 2 show unequivocally that the distribution of oxygen between water and sulphur species is primarily controlled by the pressure of degassing, and that neither the redox state of the magma source nor abundance of pre-eruptive volatiles have a significant effect. Hence, changes in the source processes such as the initiation of subduction-related volcanism should not overwhelm the fundamental pressure control on the composition of volcanic gases.

MIF records in Archaean sedimentary sulphides show that $\Delta^{33}\text{S}$ remained moderate and lower than 4‰ in the period 3.9–3.5 Gyr ago and lower than 2‰ in the period 3.5–2.7 Gyr ago^{7–9}. This must indicate volcanic emissions with low to moderate $\text{SO}_2/\text{H}_2\text{S}$ (ref. 11), a characteristic that we attribute to volcanic degassing predominantly at pressures of at least a few tens of bars (Figs 1 and 2). Conversely, the strong MIF excursion at 2.7–2.5 Gyr ago records transient volcanic emissions with high $\text{SO}_2/\text{H}_2\text{S}$ that can be obtained only through degassing at atmospheric pressure.

The Archaean tectonic style is debated. Higher mantle temperatures¹⁹ are thought to have had two effects. First, they led to the formation of thicker oceanic crust because of enhanced melting at mid-ocean ridges²⁰.

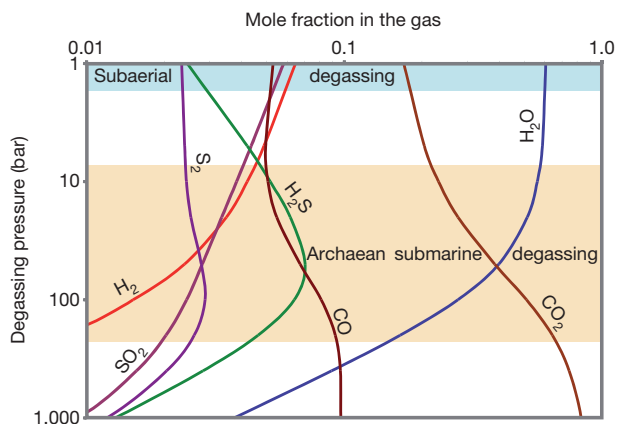


Figure 2 | Calculated compositions in mole fractions of volcanic gases as a function of pressure. The case corresponding to the black lines in Fig. 1 is shown. The inversion H_2S – SO_2 occurs at a pressure around 10 bar. Submarine degassing produces on average a $\text{SO}_2/\text{H}_2\text{S}$ ratio of 0.4, with CO_2 being the dominant gas species. Subaerial degassing yields more H_2O than CO_2 , and SO_2 is at least three times more abundant than H_2S . H_2 constitutes 6% of the gas emitted at atmospheric pressure. All calculations are performed at 1,300 °C.

variations is negligible in comparison to the effect of degassing pressure.

a, Total amount of degassed sulphur species ($\text{H}_2\text{S} + \text{S}_2 + \text{SO}_2$) expressed as weight per cent of the basaltic magma. **b**, Sulphur emitted as SO_2 expressed as weight per cent of the basaltic magma. **c**, Sulphur emitted as H_2S expressed in weight per cent of the basaltic magma. **d**, Molar $\text{SO}_2/\text{H}_2\text{S}$ ratio of volcanic gases. All calculations are performed at 1,300 °C.

Second, ocean volumes were larger because enhanced degassing in upwelling zones, and more complete and shallower dehydration of subducting oceanic crust, resulted in a larger proportion of the total water budget being at the surface^{21,22}. Higher temperatures in the continental crust (because of higher contents of radioactive heat-producing isotopes²³) reduced the strength of the continental lithosphere and resulted in subdued topography. Therefore, we attribute the decrease in venting pressure from about 100 to 1 bar at the Archaean–Proterozoic transition to a global decrease in sea level accompanying the growth of the continental crust, amplified by a decrease in ocean volume. Near-global submergence of the continents prevailed through most of the Archaean^{24,25} and almost all the volcanoes were submarine, as observed in ancient terrains²⁴. At specific times through the Precambrian era, however, the rate of crustal growth and/or continent assembly accelerated dramatically²⁶. The largest of these events, at 2.7 Gyr ago, was accompanied by a major MIF spike. We propose that this spike, and perhaps the earlier one around 3.5 Gyr ago, resulted from the transient emergence of exposed land from the global ocean. Subaerial volcanism then emitted SO_2 -dominated gas into an oxygen-poor atmosphere. The importance of such a transition in the eruptive style has been suggested before¹², but our analysis enables a quantitative description of its biogeochemical impacts. Typical compositions of volcanic gases as a function of pressure of gas venting and as a function of emerged land area are given in Supplementary Tables 1 and 2 and in Supplementary Fig. 1.

We assume that during most of the Archaean, the continents were covered by more than 100 m of water, with the consequence that volcanic gases contained little sulphur and were dominated by H_2S , with $\text{SO}_2/\text{H}_2\text{S} < 1$ (Fig. 1). Following emergence of the continents, volcanic emissions contained up to five times more sulphur and SO_2 was the dominant species with typical $\text{SO}_2/\text{H}_2\text{S}$ ratios in the range 2–9. Archaean oceans are believed to have been almost sulphate-free, in contrast to present-day sulphate-rich sea water^{27,28} (28 mmol kg^{-1}). The absence of oxygen in the Archaean atmosphere prevented oxidative alteration of continental pyrite, which is generally put forward to assert the rarity of sulphate ions in contemporaneous seas^{27,28}. However, an absence of sulphate in sea water precludes the sulphate-reduction²⁸ processes, which in turn could dramatically affect atmospheric oxygenation. Bacterial sulphate reduction produces an overall gain in atmospheric O_2 (ref. 28), which is, however, related to photosynthetic activity. But thermochemical sulphate reduction, mainly related to hydrothermal activity¹³, must have influenced oxygen balance: in the absence of thermochemical sulphate-reduction processes, O_2 provided by photosynthetic bacteria was consumed by ferrous iron dissolved in the ocean^{1,14} and by hydrothermal smokers injecting strongly reduced gases^{4,14} (Supplementary Fig. 2). Thermochemical sulphate reduction immobilized ferrous iron emitted by

hydrothermal smokers to form pyrite^{13,14}, and decreased the overall reducing capacity of hydrothermal fluids⁴. The onset of thermochemical sulphate reduction must therefore have positively affected atmospheric oxygenation (Supplementary Fig. 3).

Our calculations suggest that submarine volcanism, dominant through most of the Archaean, produced low-sulphur gases of essentially H₂S, which is much less soluble in surface waters than SO₂^{27,28}. When subaerial volcanic degassing became important in the late Archaean, it injected SO₂ that produced sulphate in sea water by disproportionation upon cooling in the atmosphere⁴ or by reduction by ferric iron in shallow waters¹¹. Consistent with recent observations²⁹, we propose that incorporation of sulphate into sea water preceded the Great Oxidation Event as subaerial volcanism became volumetrically important and then activated sulphate-reduction processes that must have resulted in O₂ enrichment of the O₂-low Archaean exosphere. Furthermore, subaerial volcanic gases are much richer in H₂ than submarine emissions (Fig. 2) and this must have enhanced hydrogen escape, a process that atmospheric models describe as an oxygenation factor for the atmosphere^{10,30}.

Figure 3 illustrates the sequence of events that may explain the relationships between MIF, volcanic gases, emergence of continents and oxygen rise during the late Archaean–early Proterozoic transition. The proposed sequence is: (1) submarine volcanism through most of the Archaean, yielding volcanic emissions with low SO₂/H₂S and moderate MIF in sediments, (2) emergence of continents in the late Archaean, which produced abundant but irregularly distributed volcanic SO₂ and led to the MIF proliferation recorded in sediments at 2.7–2.5 Gyr, (3) decomposition/oxidation of atmospheric SO₂ in sulphate, which dissolved into the oceans, (4) the start of sulphate-reduction processes, resulting in the net gain in O₂, coupled with enhanced hydrogen escape, which finally led to (5) oxygenation of the atmosphere well after the appearance of oxygenic cyanobacteria¹.

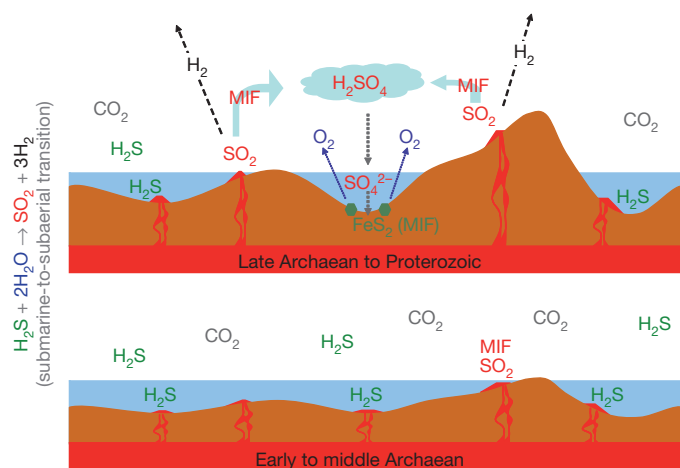


Figure 3 | Schematic illustrating the relationships between pressure of volcanic degassing and redox biogeochemical cycling. In detail, we illustrate sea level changes, continent emergence, changes in composition of volcanic gases, and the overall impact on redox biogeochemistry and oxygen rise and the Archaean–Proterozoic boundary. The lower panel shows the early to middle Archaean with large ocean volume and subdued topography, implying that most of the Earth was submerged and volcanism was predominantly submarine. Such a configuration produced volcanic gases with low SO₂/H₂S ratios, and moderate sulphur MIF. In contrast, in the late Archaean, shown in the top panel, continent emergence permitted the widespread occurrence of subaerial volcanism, characterized by gases with higher SO₂/H₂S ratios and H₂ contents. Such a configuration explains the proliferation of sulphur MIF in the period 2.7–2.5 Gyr. Increasing SO₂ contents in the atmosphere triggered an increase in seawater sulphate content, which enhanced sulphate reduction processes (Supplementary Figs 2 and 3). This, coupled with higher H₂ escape, resulted in a net gain in atmospheric O₂. The atmospheric oxygen rise, nevertheless, did not occur in the late Archaean but in the early Proterozoic.

Once oxygen was present in abundance in the atmosphere, the MIF of sedimentary sulphur disappeared.

The major finding of our study is that the sulphur MIF record of the Archaean can be explained by a change in composition of volcanic gases that facilitated oxygenation of the atmosphere. Changes in magma source processes are not the driving force; rather, the increasing SO₂/H₂S ratio of volcanic gases is explained by the increasing subaerial volcanism accompanying the emergence of continents. This, together with the rapid crustal growth that is globally observed at 2.7 Gyr, implies a change in geodynamic style that is most probably due to a cooler mantle and crust. A similar pulse of crustal growth at 3.4 Gyr ago, though smaller, may have caused the emergence of the continents and triggered the earlier MIF event.

METHODS SUMMARY

Gas–melt equilibria have been simulated using a code detailed elsewhere¹⁶. At each pressure step from 2,000 to 1 bar, we simultaneously solve the gas–melt solubility equations for H₂O–OH[−], CO₂–CO₃^{2−}, S₂–S^{2−} and the gas–gas redox equilibria involving H₂–H₂O, CH₄–CO–CO₂, H₂S–S₂–SO₂, under the requirement that the amounts of H, S, O and C remain constant in the magma (gas + melt). The iron redox state in the melt is also computed at each step. Each of these equilibrium constants has been experimentally investigated and our model combines the existing experimental data for each equilibrium into a multi-component system. Oxygen fugacity was allowed to vary in response to the various redox equilibria that occur essentially in the gas phase¹⁶, maintaining as constant the bulk oxygen/cation ratio of the system (system closed to oxygen). A system open to oxygen would not, however, change our conclusion that H₂S is dominant in the range 10–200 bar as opposed to SO₂ at atmospheric pressure. The SO₂/H₂S changes are primarily due to pressure changes, as illustrated in Figs 1 and 2. Field measurements on active volcanoes^{17,18} are consistent with our predictions. The redox state and the relative volatile abundances in the source magma are second-order parameters, implying no specific influence of the source region. We therefore assumed a common source, unlike previous studies^{11,12}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 January; accepted 16 August 2011.

1. Buick, R. When did oxygenic photosynthesis evolve? *Phil. Trans. R. Soc. Lond. B* **363**, 2731–2743 (2008).
2. Campbell, I. H. & Allen, C. M. Formation of supercontinents linked to increases in atmospheric oxygen. *Nature Geosci.* **1**, 554–558 (2008).
3. Kasting, J. F., Egger, D. H. & Raeburn, S. P. Mantle redox evolution and the oxidation state of the Archaean atmosphere. *J. Geol.* **101**, 245–257 (1993).
4. Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
5. Canil, D. Vanadium in peridotites, mantle redox and tectonic environments: Archean to present. *Earth Planet. Sci. Lett.* **195**, 75–90 (2002).
6. Li, Z. X. A. & Lee, C. T. A. The constancy of upper mantle fO₂ through time inferred from V/Sc ratios in basalts. *Earth Planet. Sci. Lett.* **228**, 483–493 (2004).
7. Farquhar, J. *et al.* Isotopic evidence for Mesoproterozoic anoxia and changing atmospheric sulphur chemistry. *Nature* **449**, 706–709 (2007).
8. Canfield, D. E., Habicht, K. S. & Thamdrup, B. The Archaean sulphur cycle and the early history of atmospheric oxygen. *Nature* **288**, 658–661 (2000).
9. Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
10. Zahnle, K. J., Claire, M. W. & Catling, D. C. The loss of mass-independent fractionation in sulfur due to a Paleoproterozoic collapse of atmospheric methane. *Geobiology* **4**, 271–283 (2006).
11. Halevy, I., Johnston, D. T. & Schrag, D. P. Explaining the structure of the Archaean mass-independent sulfur isotope record. *Science* **329**, 204–207 (2010).
12. Kump, L. R. & Barley, M. E. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* **448**, 1033–1036 (2007).
13. Butterfield, D. A. *et al.* Seafloor eruptions and evolution of hydrothermal fluid chemistry. *Phil. Trans. R. Soc. Lond. A* **355**, 369–386 (1997).
14. Kump, L. R. & Seyfried, W. E. Hydrothermal Fe fluxes during the Precambrian: effect of low oceanic sulfate concentrations and low hydrostatic pressure on the composition of black smokers. *Earth Planet. Sci. Lett.* **235**, 654–662 (2005).
15. Carmichael, I. S. E. The redox states of basic and silicic magmas: a reflection of their source regions? *Contrib. Mineral. Petrol.* **106**, 129–141 (1991).
16. Gaillard, F. & Scaillet, B. The sulfur content of volcanic gases on Mars. *Earth Planet. Sci. Lett.* **279**, 34–43 (2009).
17. Oppenheimer, C., Scaillet, B. & Martin, R. S. Sulfur degassing from volcanoes: source conditions, surveillance, plume chemistry and Earth system impacts. *Rev. Mineral. Geochem.* **73**, 363–421 (2011).
18. Aiuppa, A. *et al.* H₂S fluxes from Mt. Etna, Stromboli and Vulcano (Italy) and implications for the global volcanic sulfur budget. *Geochim. Cosmochim. Acta* **69**, 1861–1871 (2005).

19. Bickle, M. J. Heat loss from the Earth: a constraint on Archaean tectonics from the relation between geothermal gradients and the rate of heat production. *Earth Planet. Sci. Lett.* **40**, 301–315 (1978).
20. Sleep, N. H. & Windley, B. F. Archaean plate tectonics—constraints and inferences. *J. Geol.* **90**, 363–379 (1982).
21. Kasting, J. F. & Holm, N. G. What determines the volume of the oceans. *Earth Planet. Sci. Lett.* **109**, 507–515 (1992).
22. Bounama, C., Franck, S. & von Bloh, W. The fate of the Earth's ocean. *Hydrol. Earth Syst. Sci.* **5**, 569–575 (2001).
23. Sandiford, M. & McLaren, S. in *Evolution and Differentiation of the Continental Crust* (eds Brown, M. & Rushmer, T) 67–92 (Cambridge University Press, 2006).
24. Arndt, N. T. Why was flood volcanism on submerged continental platforms so common in the Precambrian? *Precamb. Res.* **97**, 155–164 (1998).
25. Flament, N., Coltice, N. & Rey, P. F. A case for late-Archaean continental emergence from thermal evolution models and hypsometry. *Earth Planet. Sci. Lett.* **275**, 326–336 (2008).
26. Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* 1–312 (Blackwell, 1985).
27. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archaean ocean. *Science* **298**, 2372–2374 (2002).
28. Lyons, T. W. & Gill, B. C. Ancient sulfur cycling and oxygenation of the early biosphere. *Elements* **6**, 93–99 (2010).
29. Scott, C. T. *et al.* Late Archean euxinic conditions before the rise of atmospheric oxygen. *Geology* **39**, 119–122 (2011).
30. Claire, M. W., Catling, D. C. & Zahnle, K. J. Biogeochemical modelling of the rise in atmospheric oxygen. *Geobiology* **4**, 239–269 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Lyons, D. Canil and N. Sleep for comments. We acknowledge support by INSU-PNP and the ANR (grants ANR-10-BLAN-60301 and ANR-10-BLAN-62101).

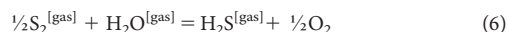
Author Contributions All authors contributed to discussions and writing. F.G. performed the calculations and wrote the first draft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.G. (gaillard@cnrs-orleans.fr).

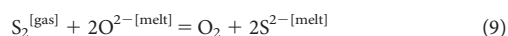
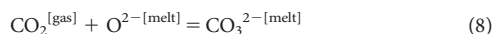
METHODS

Following ref. 16, the mole fraction of the different gas species is calculated from the thermodynamic constants of the reactions occurring in the system C–H–S–O at magmatic pressure and temperature³¹.

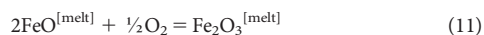
The homogeneous equilibria in the gas phase are



In addition, the following melt–gas equilibria are taken into account using thermodynamic data from ref. 32 for water and carbon dioxide, and ref. 33 for sulphur and ref. 34 for H₂:



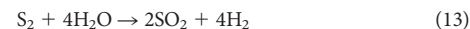
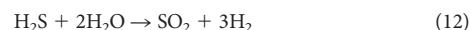
The iron redox ratio in the melt is computed following ref. 35 as



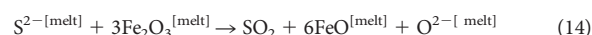
Each of the equilibrium constants (reactions (2) to (10)) has been combined in the multi-component C–O–H–S system. The overall uncertainties in calculated gas compositions are 10%–15%. These uncertainties are mainly inherited from the experimental calibration of C–O–H–S species solubility^{31–34}. During degassing, the bulk mass fractions of the different volatile components (H, S, C, O) are kept constant and gas–melt, gas–gas and melt–melt equilibria are simultaneously solved to derive the gas species abundances at each pressure step (from 2,000 to 1 bar). The number of pressure steps imposed in the calculations has no effect on the results because thermodynamic equilibrium does not depend on the pressure–temperature path but only on the final pressure and temperature considered. The initial abundance of C–O–H–S components is fixed at 2,000 bar, which sets the

amount of dissolved volatiles and both quantity and composition of the coexisting gas fraction. The system was computed as a system closed to oxygen^{16,36} but the differences between open or closed system conditions remain negligible.

Upon decompression, because volatile solubilities in the melt decreases, the gas fraction increases and its composition changes, being enriched in highly volatile species (CO₂) at high pressure and progressively diluted by less volatile species (H₂O and sulphur species) with decreasing pressure. Sulphur species display a critical change at low pressure: SO₂, nearly absent at elevated degassing pressures, dominates at 1 bar over H₂S and S₂. The stoichiometry of those changes follows the reactions



for gas–gas equilibria and the reaction



for melt–gas equilibria.

In reactions (12) and (13), the right-hand side involves a higher mole number and is therefore favoured by decreasing pressure. In reaction (14), sulphur degassing as SO₂ at low pressure implies a reduction of ferric into ferrous iron, consistent with geochemical observations¹⁵. The importance of reaction (14), which may induce a decrease in melt oxygen fugacity, is, however, limited by the kinetics of melt reduction³⁷. This kinetics delay between reactions (12) and (13) and reaction (14), which we do not take into account, allows the gas SO₂/H₂S ratio to be higher than predicted in Figs 1 and 2, potentially exceeding 10 at 1 bar. The thermochemistry of high-temperature gas–melt volcanic systems therefore offer the most likely explanation of variation in sulphur MIF observed during the Archean.

31. Shi, P. F. & Saxena, S. K. Thermodynamic modelling of the C–H–O–S fluid system. *Am. Mineral.* **77**, 1038–1049 (1992).
32. Morizet, Y., Paris, M., Gaillard, F. & Scaillet, B. C–O–H fluid solubility in haplobasalt under reducing conditions: an experimental study. *Chem. Geol.* **279**, 1–16 (2010).
33. O'Neill, H. S. C. & Mavrogenes, J. The sulfide saturation capacity and the sulphur content at sulfide saturation of silicate melts at 1400 °C and 1 bar. *J. Petrol.* **43**, 1049–1087 (2002).
34. Gaillard, F., Schmidt, B. C., Mackwell, S. & McCammon, C. Rate of hydrogen-iron redox exchange in silicate melts and glasses. *Geochim. Cosmochim. Acta* **67**, 2427–2441 (2003).
35. Kress, V. C. & Carmichael, I. S. E. The compressibility of silicate liquids containing Fe₂O₃ and the effect of composition, temperature, oxygen fugacity and pressure on their redox states. *Contrib. Mineral. Petrol.* **108**, 82–92 (1991).
36. Burgisser, A. & Scaillet, B. Redox evolution of degassing magma rising to the surface. *Nature* **445**, 194–197 (2007).
37. Pommier, A., Gaillard, F. & Pichavant, M. Time-dependent changes of the electrical conductivity of basaltic melts with redox state. *Geochim. Cosmochim. Acta* **74**, 1653–1671 (2010).

Strong contributors to network persistence are the most vulnerable to extinction

Serguei Saavedra^{1,2,3*}, Daniel B. Stouffer^{4,5*}, Brian Uzzi^{1,2} & Jordi Bascompte⁴

The architecture of mutualistic networks facilitates coexistence of individual participants by minimizing competition relative to facilitation^{1,2}. However, it is not known whether this benefit is received by each participant node in proportion to its overall contribution to network persistence. This issue is critical to understanding the trade-offs faced by individual nodes in a network^{3–5}. We address this question by applying a suite of structural and dynamic methods to an ensemble of flowering plant/insect pollinator networks. Here we report two main results. First, nodes contribute heterogeneously to the overall nested architecture of the network. From simulations, we confirm that the removal of a strong contributor tends to decrease overall network persistence more than the removal of a weak contributor. Second, strong contributors to collective persistence do not gain individual survival benefits but are in fact the nodes most vulnerable to extinction. We explore the generality of these results to other cooperative networks by analysing a 15-year time series of the interactions between designer and contractor firms in the New York City garment industry. As with the ecological networks, a firm's survival probability decreases as its individual nestedness contribution increases. Our results, therefore, introduce a new paradox into the study of the persistence of cooperative networks, and potentially address questions about the impact of invasive species in ecological systems and new competitors in economic systems.

Mutualistic interactions form the basis of many biological and human systems of cooperation and competition^{2,6–14}. Mutualistic networks are composed of mutually beneficial interactions between individual participants or nodes of two distinct sets, such as plant species and their pollinators¹⁵ or designers and their contractors¹¹. One pattern in particular—nestedness—appears ubiquitous in mutualistic networks from a variety of contexts^{1,2,16}. In a nested network, the interactions are organized such that specialists (for example, plants with few pollinators) interact with proper subsets of the nodes with whom generalists (for example, plants with many pollinators) interact¹⁶. This nested architecture has been shown to minimize competition between species and therefore allows the network to support greater biodiversity¹.

Although greater nestedness allows for the successful coexistence of more species, it is unclear how the decreased risk of extinction is distributed among the nodes in the network. Here we quantify whether node-level survival benefits are related to each node's 'contribution' to the nested architecture, defined as the degree to which the organization of their interactions increases overall nestedness. A positive relationship could create a positive feedback of benefits to those that most support nestedness and the community; a negative relationship, in contrast, could imply that some nodes stand to benefit from the contributions of others.

To answer these questions, we integrate structural and dynamic analyses of 20 ecological networks that describe mutually beneficial interactions between flowering plants and their insect pollinators across

diverse environmental and biotic conditions¹⁷. In these bipartite networks, nodes correspond to individual plant or pollinator species and links between nodes indicate that a pollinator species has been found empirically to pollinate a given plant species (Methods).

To measure the individual contribution to nestedness for each species or node, we develop a novel, node-level metric that quantifies how an individual's contribution to network nestedness compares to that expected at random (Fig. 1). The measure quantifies the degree to which the overall nestedness of the network compares with the value obtained when randomizing just the interactions of that particular node.

Mathematically, this is defined as $c_i = (N - \langle N_i^* \rangle) / \sigma_{N_i^*}$, where N is the observed nestedness of the network and $\langle N_i^* \rangle$ and $\sigma_{N_i^*}$ are the average and standard deviation of nestedness across an ensemble of random replicates within which the interactions of node i have been randomized (Methods). The greater the degree to which the interactions of node i are consistent with the network's overall nestedness, the stronger is this node's contribution c_i , and vice versa.

We calculate this measure for each species in each of the networks and observe that node contributions to the network architecture are

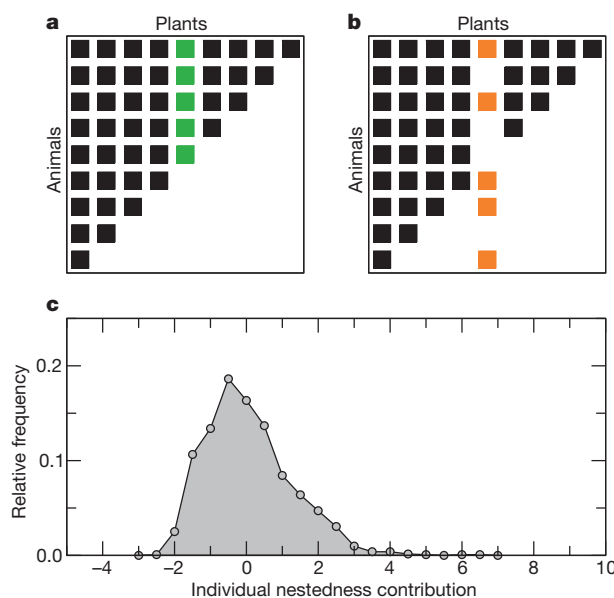


Figure 1 | Nodes contribute to the nested architecture of the network in distinct proportions. **a, b,** The individual nestedness contribution of a node (for example, the plant species whose interactions are highlighted in green in **a**) is defined as the degree to which the observed network nestedness compares to the value obtained when randomizing just the interactions of that particular node, highlighted in orange in **b**. **c,** The empirical distribution of individual nestedness contribution for all species in the 20 pollination networks studied here.

¹Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA. ²Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, USA.

³Northwestern University Clinical and Translational Sciences Institute, Northwestern University, Chicago, Illinois 60611, USA. ⁴Integrative Ecology Group, Estación Biológica de Doñana, CSIC, Calle Américo Vespucio s/n, E-41092 Sevilla, Spain. ⁵School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand.

*These authors contributed equally to this work.

heterogeneously distributed across our empirical data set (Fig. 1c). Importantly, some species contribute comparatively little to the nested structure of the network whereas others contribute considerably more.

Having shown that species contribute to the overall network architecture in distinct proportions, we next explore the systemic consequences of species extinctions across the spectrum of contributions. This allows us to test whether or not a node's contribution in topological terms translates to a dynamic contribution in terms of network persistence. As there are no dynamic, empirical data with which to quantify network persistence in our ecological networks, we simulate species dynamics with a recently published model that is appropriate for mutualistic systems¹ (Methods). To measure the dynamic impact of a node on overall network persistence, we calculate the difference between the persistence of the network with and without removal of the focal node. Network persistence is measured as the fraction of initial species remaining at the end of the simulation. Here, we consider plant and pollinator species pooled together, while in Supplementary Information we calculate the persistence for each set independently.

We find that the more a node contributes to nestedness, the more likely it is that its loss is detrimental to the network's persistence (Fig. 2). Nestedness contribution therefore represents a key measure of the degree to which each node's interactions work for or against the long-term persistence of species in the mutualistic network.

Because the extinction of strong contributors has significant repercussions on network persistence, we proceed by estimating the vulnerability of these strong contributors to extinction. Specifically, we compare each node's nestedness contribution to its survival probability, where survival is determined by whether or not the node goes extinct before the dynamic simulations reach equilibrium.

Surprisingly, we find that nodes that contribute the most to the nestedness of the network—and its persistence—are the most likely to go extinct (Fig. 3). Indeed, individual nestedness contribution has a significant, negative correlation with survival probability (Methods). This conclusion is independent of whether all networks are analysed together ($P < 10^{-8}$ and $P < 10^{-15}$ for plants and pollinators, respectively) or each network is analysed separately. Specifically, the negative relationship between contribution and survival probability is significant for pollinators in 17 out of 20 networks and for plants in 20 out of 20 networks ($P < 10^{-4}$). Furthermore, our results capture the importance of nestedness contribution above and beyond the effect of the number of interactions per node (Methods). In general, the more a node contributes to the architecture of its network, the greater its

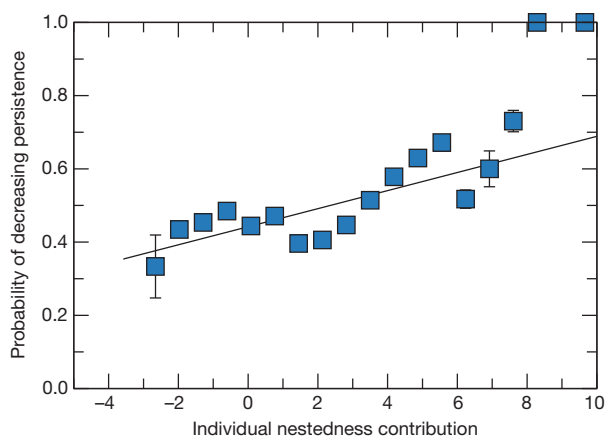


Figure 2 | The extinction of stronger contributors leads to a decrease in network persistence. We plot the probability that a node's removal causes a decrease in network persistence in the dynamic simulations as a function of that node's contribution to nestedness. All species in the 20 pollination networks are plotted together. Error bars, standard errors of the reported averages; in some cases they are smaller than the plotting symbols. Solid line, best-fit linear regression ($P < 0.005$).

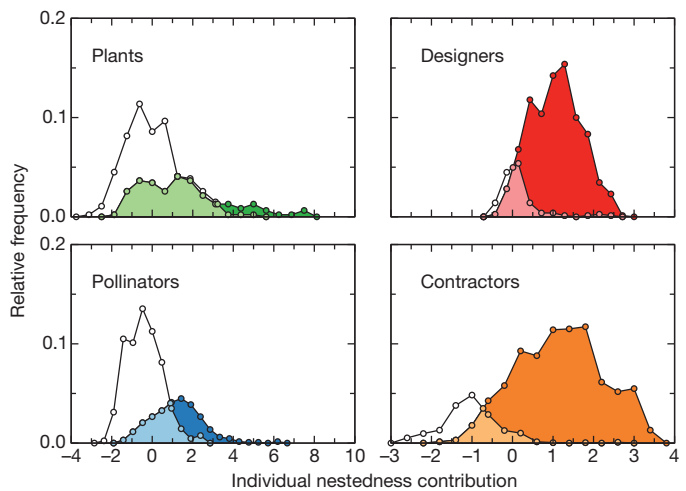


Figure 3 | Strong contributors to nestedness are the most vulnerable to extinction. We plot histograms of individual nestedness contribution for nodes that survive, with the area under the curve shaded in white, and for those that do not, with the area under the curve shaded in colour. In each case, the lighter colour indicates overlap between the two distributions. Nodes from all networks are pooled together for each class of node studied.

probability of extinction. These results imply that some species can benefit from others by participating in mutualistic interactions that differ from those that would maximize network persistence.

To explore the generality of these results to other types of cooperative networks, we build on previous work that illustrates commonalities between ecological and human systems^{2,9,11}. Specifically, we apply our node-contribution measure to the network formed by the cooperative interactions between designers and their contractors in the New York City garment industry across a 15-year observation period¹⁸. This industry is characterized by a competitive and dynamic environment where resource exchanges among firms and survival depends on collaborative links between firms^{11,18}. Though earlier studies have attributed the failure of firms to their lack of adaptation to new business networks^{19–21}, none have unambiguously linked firm survival to the network architecture. Here, nodes correspond to an individual designer or contractor firm, and links between nodes indicate that a designer exchanged money for the contractor's production services. We quantify a firm's survival by comparing whether or not the company was still operating at the end of the observation period.

Importantly, the negative relationship that we observed in ecological networks—between a node's contribution and its survival—also holds for the firms in this socio-economic network. This conclusion is obtained both when looking at the companies that went out of business by the end of the 15-year period ($P < 10^{-11}$ and $P < 10^{-12}$ for designers and contractors, respectively) and when tracking the yearly dynamics across the data set. In 14 out of 15 year-to-year intervals, the relationship between individual nestedness contribution and node survival is significant and negative for both contractors and designers ($P < 10^{-4}$). Just as for nodes in the ecological networks, the nodes that contribute the most to the nestedness of the network are the most likely to have failed (Fig. 3). This analysis suggests that our results may apply generally across different types of cooperative networks, whether they are shaped as a consequence of evolution or conscious decisions.

The individual nestedness contribution that we define here provides a means to estimate the expected survival of participants in cooperative systems solely from knowledge of the network structure. Moreover, we have revealed a paradox of nestedness: although strong contributors to nestedness are more important for the persistence of the entire network, they are also more prone to extinction compared to those nodes that contribute proportionally less. Although there is no clear explanation for this result, one could speculate that nodes that conform to the

architectural expectation of a nested network are subject to greater constraints than nodes that interact freely. If these linkage constraints have a cost in terms of fitness (for example, they lead to a pollinator species interacting with a less rewarding plant species), they could ultimately translate to a higher probability of extinction. Our study therefore raises new questions about the origins of nodes that make strong contributions to the collective good and nodes that appear to improve their own survival at the expense of others^{4,5,14}. In ecology, our results could inform a quantitative assessment of the likely persistence of an invasive species in the network and its effects on the overall welfare of the community. In socio-economic systems, our results could be used to identify those companies and economic sectors that undermine stable, long-term economic prosperity, as well as to develop interventions that take into account collective interests.

METHODS SUMMARY

Our ecological data set contains the 20 largest plant–pollinator mutualistic networks provided in the Supplementary Information of ref. 17; details about the networks and original sources can be found therein. Additionally, we analyse the network between designer and contractor firms in the New York City garment industry¹¹. Details about this temporal network can be found in the Supplementary Information of ref. 11.

Nestedness N is quantified using the measure proposed in ref. 22. In calculating nestedness contributions, the interactions of a node were randomized according to the null model specified in ref. 16; we used 1,000 random replicates. Note that all results presented in the present Letter hold both for alternative measures of nestedness and alternative null models (Supplementary Methods; Supplementary Figs 1–6).

To explore network dynamics, we employed the mutualistic model defined in ref. 1. This model is based on a system of differential equations describing the dynamics of P plant species and A animal species as a function of their intrinsic growth rates, interspecific competition, and mutualistic effects of one set on another.

We quantified the relationship between nestedness contribution c_i and node's survival probability s_i via a logistic regression. To ensure that our analysis is capturing a significant pattern over and above a node's degree (that is, its number of interactions), two additional terms in the regression controlled for the degree of the node and the potential interaction between degree and contribution to nestedness.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 April; accepted 5 August 2011.

Published online 14 September 2011.

1. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458**, 1018–1020 (2009).
2. Sugihara, G. & Ye, H. Cooperative network dynamics. *Nature* **458**, 979–980 (2009).
3. Rankin, D. J., Bargum, K. & Kokko, H. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**, 643–651 (2007).
4. Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proc. Natl Acad. Sci. USA* **74**, 4542–4546 (1977).

5. Falster, D. S. & Westoby, M. Plant height and evolutionary games. *Trends Ecol. Evol.* **18**, 337–343 (2003).
6. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
7. Gintis, H., Bowles, S., Boyd, R. T. & Fehr, E. (eds) *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (MIT Press, 2005).
8. Bascompte, J. Disentangling the web of life. *Science* **325**, 416–419 (2009).
9. May, R. M., Levin, S. A. & Sugihara, G. Complex systems: ecology for bankers. *Nature* **451**, 893–895 (2009).
10. Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* **469**, 351–355 (2011).
11. Saavedra, S., Reed-Tsochas, F. & Uzzi, B. A simple model of bipartite cooperation for ecological and organisational networks. *Nature* **457**, 463–466 (2009).
12. Saavedra, S., Powers, S., McCotter, T., Porter, M. A. & Mucha, P. J. Mutually-antagonistic interactions in baseball networks. *Physica A* **389**, 1131–1141 (2010).
13. Fowler, J. H. & Christakis, N. Cooperative behavior cascades in human social networks. *Proc. Natl Acad. Sci. USA* **107**, 5334–5338 (2010).
14. Weyl, E. G., Frederickson, M. E., Yu, D. W. & Pierce, N. E. Economic contract theory tests models of mutualism. *Proc. Natl Acad. Sci. USA* **107**, 15712–15716 (2010).
15. Bascompte, J. & Jordano, P. The structure of plant–animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Syst.* **38**, 567–593 (2007).
16. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
17. Rezendes, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P. & Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**, 925–928 (2007).
18. Saavedra, S., Reed-Tsochas, F. & Uzzi, B. Asymmetric disassembly and robustness in declining networks. *Proc. Natl Acad. Sci. USA* **10**, 16466–16471 (2009).
19. De Toni, A. & Nassimbeni, G. Supply networks: genesis, stability and logistics implications. A comparative analysis of two districts. *Omega* **23**, 403–418 (1995).
20. Uzzi, B. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. *Am. Sociol. Rev.* **61**, 674–698 (1996).
21. Doeringer, J. & Crean, S. Can fast fashion save the US apparel industry? *Socioecon. Rev.* **4**, 353–377 (1996).
22. Almeida-Neto, M., Guimarães, P., Guimarães, P. R. Jr, Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227–1239 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding was provided by the Kellogg School of Management, Northwestern University, the Northwestern University Institute on Complex Systems (NICO; to S.S. and B.U.), NUCATS grant UL1RR025741 (to S.S.), a CSIC-JAE postdoctoral fellowship (to D.B.S.), the Army Research Laboratory (under cooperative agreement W911NF-09-2-0053 to B.U.), and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) through an Advanced Grant (grant agreement 268543 to J.B.). Figures were generated with PyGrace (<http://pygrace.sourceforge.net>).

Author Contributions S.S. and D.B.S. analysed the data and performed the simulations. S.S., D.B.S., B.U. and J.B. designed the study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.B. (bascompte@ebd.csic.es).

METHODS

Data sets. Here we analyse two types of empirical data. First, we analyse a data set containing the 20 largest plant–pollinator mutualistic networks provided in the Supplementary Information of ref. 17. Further details about the network size, species composition and geographic location can be found in ref. 17, together with the actual network and original source. Additionally, we analyse the socio-economic network first described in ref. 11. This network contains approximately 700,000 commercial interactions between manufacturer and contractor firms in the New York City garment industry between January 1985 and December 2003. From this data set, we can generate yearly snapshots of the bipartite network. Additional details can be found in the Supplementary Information of ref. 11.

Nestedness measure. We quantify nestedness using NODF, the measure recently introduced in ref. 22. This measure reduces potential bias introduced by network size and shape compared with alternative measures. The overall nestedness of the network N is defined mathematically as:

$$N = \frac{\sum_{i < j}^P M_{ij} + \sum_{i < j}^A M_{ij}}{\left[\frac{P(P-1)}{2} \right] + \left[\frac{A(A-1)}{2} \right]}$$

where the first sum is across all pairs of plant species, the second sum is across all pairs of animal species, and P and A are the total number of plant species and animal species, respectively. For every pair of nodes i and j , $M_{ij} = 0$ if $k_i = k_j$, and $M_{ij} = n_{ij}/\min(k_i, k_j)$ otherwise. Here, k is a node's number of interactions; n_{ij} is the number of interactions in common between nodes i and j ; and $\min(k_i, k_j)$ refers to the minimum of the two values k_i and k_j .

This nestedness metric takes values in the interval $N \in [0, 1]$, where 1 designates a perfectly nested network and 0 indicates a network with no nestedness. Alternative measures of nestedness are highly and significantly correlated to this one (Supplementary Methods; Supplementary Fig. 1). Furthermore, all results presented in this Letter hold for such alternative measures of nestedness (Supplementary Figs 2 and 3).

Null model. To randomize interactions, we use the null model outlined in ref. 16. Under the specifications of this null model, the interactions are assigned according to the rule:

$$p_{ij} = \frac{1}{2} \left(\frac{k_i}{P} + \frac{k_j}{A} \right)$$

where p_{ij} is the probability of an interaction between node i (of set A) and node j (of set P) and A and P are the total number of animal and plant species in sets A and P ,

respectively. In the socio-economic networks, the model is specified in the same fashion except with designers and contractors in the place of animals and plants. We use 1,000 replicates. Note that all results presented in this Letter hold for alternative null models (Supplementary Methods; Supplementary Figs 4, 5, and 6).

Dynamic model for ecological mutualistic networks. To simulate interspecies dynamics in the ecological networks, we run the dynamic model of ref. 1 using each of the real networks as the skeleton of the model. The real networks specify the number of plant species, the number of animal species, and who interacts with whom. In the dynamic model, the change in abundance over time for a plant species i follows:

$$\frac{dS_i^{(P)}}{dt} = \alpha_i^{(P)} S_i^{(P)} - \sum_{j \in P} \beta_{ij}^{(P)} S_i^{(P)} S_j^{(P)} + \sum_{k \in A} \frac{\gamma_{ik}^{(P)} S_i^{(P)} S_k^{(A)}}{1 + h^{(P)} \sum_{l \in A} \gamma_{il}^{(P)} S_l^{(A)}}.$$

The same equations for animal species can be written in a symmetric form interchanging the indices (P) and (A) .

To fully specify the remainder of the dynamic model, we use the following parameter values for all plant species (P) and animal species (A): intrinsic growth rates α_i are drawn uniformly from the interval $[0.85, 1.1]$; the competitive interactions β_{ii} and β_{ij} are drawn uniformly from the intervals $[0.99, 1.01]$ and $[0.22, 0.24]$, respectively; the mutualistic interactions γ_{ij} , encapsulating the *per capita* effect of animal j on plant i , are drawn uniformly from the interval $[0.19, 0.21]$; the handling time h is set to 0.1.

Simulations are performed by integrating the system of ordinary differential equations using a fourth-order Runge-Kutta method with small integration steps. All initial abundance densities S_i are drawn uniformly from the interval $(0, 1]$. Species are considered to have gone extinct when their abundance density S_i is lower than 10^{-30} . All results are robust to changes in parameter values (growth rates $\alpha_i \in [0, 2]$, competitive interactions $\beta_{ij} \in [0, 1]$, and mutualistic interactions $\gamma_{ij} \in [0, 1]$) as well as changing the functional responses of the above model from Holling type II to Holling type III.

Relationship between nestedness contribution and node survival. We quantify the relationship between nestedness contribution c_i and node's survival probability s_i by using a logistic regression with the form $\text{logit}(s_i) = \alpha + \beta c_i$. Survival was coded as 0 and 1 for non-surviving and surviving nodes, respectively. To ensure that our analysis is capturing a significant pattern above and beyond other network attributes, we perform the same analysis but also include terms for node degree k (that is, number of interactions) and the potential interaction term for degree and contribution. This extended model takes the form $\text{logit}(s_i) = \alpha + \beta c_i + \gamma k_i + \delta c_i k_i$.

An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*

Yael Grosjean^{1,2}, Raphael Rytz¹, Jean-Pierre Farine², Liliane Abuin¹, Jérôme Cortot², Gregory S. X. E. Jefferis³ & Richard Benton¹

Many animals attract mating partners through the release of volatile sex pheromones, which can convey information on the species, gender and receptivity of the sender to induce innate courtship and mating behaviours by the receiver¹. Male *Drosophila melanogaster* fruitflies display stereotyped reproductive behaviours towards females, and these behaviours are controlled by the neural circuitry expressing male-specific isoforms of the transcription factor Fruitless (*FRU^M*)^{2–5}. However, the volatile pheromone ligands, receptors and olfactory sensory neurons (OSNs) that promote male courtship have not been identified in this important model organism. Here we describe a novel courtship function of Ionotropic receptor 84a (IR84a), which is a member of the chemosensory ionotropic glutamate receptor family⁶, in a previously uncharacterized population of *FRU^M*-positive OSNs. IR84a-expressing neurons are activated not by fly-derived chemicals but by the aromatic odours phenylacetic acid and phenylacetaldehyde, which are widely found in fruit and other plant tissues⁷ that serve as food sources and oviposition sites for drosophilid flies⁸. Mutation of *Ir84a* abolishes both odour-evoked and spontaneous electrophysiological activity in these neurons and markedly reduces male courtship behaviour. Conversely, male courtship is increased—in an IR84a-dependent manner—in the presence of phenylacetic acid but not in the presence of another fruit odour that does not activate IR84a. Interneurons downstream of IR84a-expressing OSNs innervate a pheromone-processing centre in the brain. Whereas IR84a orthologues and phenylacetic-acid-responsive neurons are present in diverse drosophilid species, IR84a is absent from insects that rely on long-range sex pheromones. Our results suggest a model in which IR84a couples food presence to the activation of the *fru^M* courtship circuitry in fruitflies. These findings reveal an unusual but effective evolutionary solution to coordinate feeding and oviposition site selection with reproductive behaviours through a specific sensory pathway.

While mapping the projections of Ionotropic receptor (IR)-expressing OSNs to the primary olfactory centre⁹, the antennal lobe, we observed that an *Ir84a* reporter was labelling neurons innervating the VL2a glomerulus. VL2a is one of only three glomeruli that are larger in males and whose OSN inputs and projection neuron outputs express male-specific isoforms of the behavioural sex determination gene *fruitless* (*fru^M*)^{4,5}. *fru^M*-expressing OSNs have been implicated in promoting male sexual behaviours, because inhibition of synaptic transmission in all of these neurons simultaneously reduces male courtship of females⁵. We confirmed the expression of *Ir84a* in *fru^M*-expressing neurons by visualizing the co-expression of an *Ir84a* reporter, as well as endogenous *Ir84a* transcripts, with a *fru^M* reporter^{4,5} (Fig. 1a and Supplementary Fig. 1a). No sexual dimorphism was observed either in the number of *Ir84a*-expressing cells (males, 30 ± 2 neurons; females, 30 ± 3 neurons; $n = 4$ antennae each) or in their targeting to VL2a, indicating that *FRU^M* does not have an essential role in the development of these neurons, similar to other *fru^M*-expressing OSNs⁵.

We generated a *GAL4* knock-in null allele, *Ir84a^{GAL4}* (Fig. 1b, c). *Ir84a^{GAL4/+}* heterozygotes expressed a *GAL4*-responsive, membrane-targeted, green fluorescent protein (GFP) transgene (*UAS-mCD8:GFP*) exclusively in *Ir84a*-expressing OSNs (Fig. 1d, *Ir84a^{+/-}*, and Supplementary Fig. 1b). In *Ir84a^{GAL4}* homozygotes, the endogenous expression of *Ir84a* was lost, but the distribution and dendritic projections of these neurons, as revealed by mCD8:GFP, was unaffected (Fig. 1d, *Ir84a^{-/-}*). The axons of *Ir84a*-expressing neurons in heterozygous and homozygous *Ir84a^{GAL4}* flies projected only to VL2a (Fig. 1e). *Ir84a* is therefore dispensable for the specification and wiring of the neurons in which it is expressed. An amino-terminal enhanced GFP (EGFP)-tagged version of this receptor¹⁰ localized to the cell bodies and the ciliated dendritic endings of these neurons but not to their axon termini (Fig. 1f, g), consistent with an exclusive role for IR84a as an olfactory receptor in the *fru^M* circuitry.

We tested the responses of IR84a-expressing neurons to chemicals produced by male or virgin female flies¹¹, both by delivering headspaces of flies from a distance (simulating the action of volatile pheromones) and by presenting extracts from fly cuticles at close range (mimicking exposure to non-volatile hydrocarbons, such as contact pheromones¹²). These stimuli produced no or extremely small responses, as detected by extracellular recordings in ac4 sensilla (Supplementary Fig. 2a), which belong to the class of olfactory hair that houses IR84a-expressing neurons, as well as OSNs that express IR75d, or IR76a and IR76b⁶. These observations suggest that IR84a is not tuned to fly-derived pheromones. We therefore tested 163 structurally diverse odours. Only three of these gave responses of >50 spikes s^{-1} above basal activity: phenylacetaldehyde (as identified previously¹³), phenylacetic acid and phenylethylamine (Fig. 2a and Supplementary Fig. 2b). Dose response curves that revealed sensitivity to these ligands are similar in both sexes (Fig. 2b and data not shown).

In *Ir84a^{GAL4}* homozygous mutants, the responses to phenylacetic acid and phenylacetaldehyde were completely abolished (Fig. 2c, d, *Ir84a^{-/-}*). Re-introduction of *Ir84a* function in these neurons, by using *UAS-Ir84a* or *UAS-EGFP:Ir84a* cDNA transgenes, rescued these phenotypes (Fig. 2c, d, rescue), indicating a cell-autonomous function for IR84a in mediating these odour responses. By contrast, responses to phenylethylamine were unaffected (Fig. 2c, d), corroborating the evidence that this chemical is detected by the neurons that express both IR76a and IR76b^{9,10}. Consistent with these loss-of-function data, mis-expression of IR84a in Odorant receptor 35a (OR35a)-expressing neurons was sufficient to confer responsiveness to phenylacetic acid and phenylacetaldehyde (Supplementary Fig. 2c). The basal activity in *Ir84a* mutant ac4 sensilla was also lower than that in the ac4 sensilla of wild-type and rescue genotypes (Fig. 2e, f), indicating that IR84a has a role in promoting spontaneous firing.

Phenylacetic acid and phenylacetaldehyde are aromatic compounds found in a diverse range of fruit and other plant tissues⁷, as well as in their fermentation products¹⁴, and they are used in human perfumes for their floral, honey-like, sweet smell. We confirmed the presence of these chemicals in two host fruit for drosophilid flies⁸, overripe bananas and

¹Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland. ²Centre des Sciences du Goût et de l'Alimentation, UMR-6265 CNRS, UMR-1324 INRA, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France. ³Division of Neurobiology, MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.

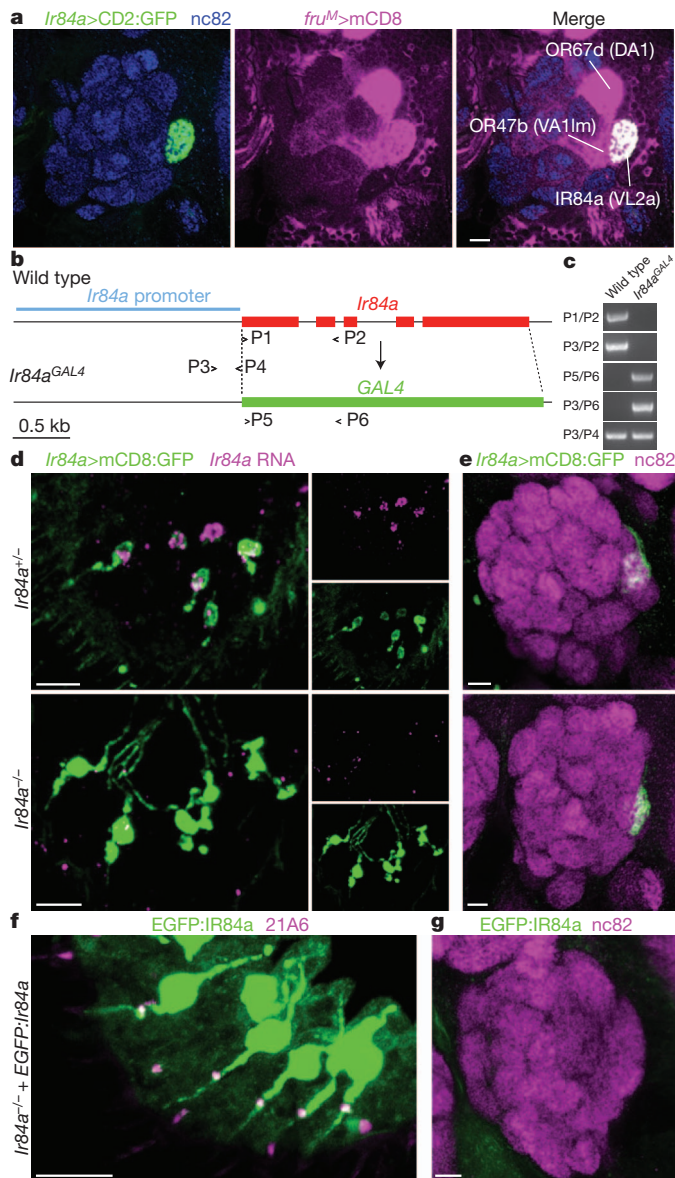


Figure 1 | Gene targeting of *Ir84a*, a candidate olfactory receptor in the *fru^M* circuit. **a**, Immunofluorescence using anti-GFP (green), anti-CD8 (magenta) and nc82 (blue; neuropil marker) antibodies on an antennal lobe of an animal in which an *Ir84a* promoter-*lexA* transgene drives (>) expression of a LexA-responsive CD2:GFP reporter transgene and in which the *fru^M* GAL4 insertion allele drives expression of a GAL4-responsive CD8 reporter transgene (genotype, *Ir84a-lexA/UAS-mCD8;lexAop-rCD2:GFP/fru^MGAL4*). **b**, Schematic of the *Ir84a* locus, illustrating the organization of wild-type *Ir84a* and null mutant *Ir84a^{GAL4}* alleles. The span of the cloned *Ir84a* promoter is also indicated. P denotes PCR primer position. kb, kilobases. **c**, PCR confirmation of the *Ir84a^{GAL4}* allele using the pairs of primers shown in **b**. **d**, Combined *Ir84a* RNA *in situ* hybridization (magenta) and anti-GFP immunofluorescence (green) on antennae from *Ir84a* mutant flies: top, heterozygous mutants (*UAS-mCD8:GFP/+;Ir84a^{GAL4}/+*); and, bottom, homozygous mutants (*UAS-mCD8:GFP/+;Ir84a^{GAL4}/Ir84a^{GAL4}*). Separate channels are shown on the right. **e**, Immunofluorescence using anti-GFP (green) and nc82 (magenta) antibodies on antennal lobes from *Ir84a* mutant flies: top, control heterozygous mutants (*UAS-mCD8:GFP;Ir84a^{GAL4}/+*); and, bottom, homozygous mutants (*UAS-mCD8:GFP;Ir84a^{GAL4}/Ir84a^{GAL4}*). **f**, Immunofluorescence using anti-GFP antibodies (green) and antibodies specific for the sensory cilium base marker 21A6 (magenta) on an antenna expressing EGFP:Ir84a (*UAS-EGFP:Ir84a/+;Ir84a^{GAL4}/Ir84a^{GAL4}*). **g**, Immunofluorescence using anti-GFP (green) and nc82 (magenta) antibodies on an antennal lobe of a fly expressing EGFP:Ir84a (*UAS-EGFP:Ir84a;Ir84a^{GAL4}/Ir84a^{GAL4}*). **a**, **d**–**g**, Scale bar, 10 μ m.

the prickly-pear cactus *Opuntia ficus-indica*, as well as in laboratory *Drosophila* medium, by using gas chromatography–mass spectrometry analysis (Supplementary Table 1). The ubiquity of phenylacetic acid in vegetal tissues may be linked with its activity as a growth-regulating auxin⁷ and/or its production by plant-associated microorganisms¹⁵. Small, but reproducible, quantities of phenylacetic acid and phenylacetaldehyde were also detected in whole-body cuticular extracts of male and virgin female *D. melanogaster* (Supplementary Table 1). The similarity in the relative amounts of these chemicals in laboratory medium and fruitfly extracts (Supplementary Table 1) suggested that these chemicals are transferred from food to flies during their culture. ‘Clean’ cuticular extracts from animals grown on a minimal medium containing only sucrose and agarose consistently contain no detectable phenylacetaldehyde or phenylacetic acid (Supplementary Table 1).

The expression of IR84a in *fru^M*-expressing neurons implicates this receptor in the regulation of male courtship^{4,5}. Indeed, in single-pair courtship assays, we observed that *Ir84a^{GAL4}* mutant males court wild-type females significantly less than do wild-type males (Fig. 3a and Supplementary Fig. 3a, b). This phenotype was observed using both decapitated virgin females (which do not produce feedback signals) (Fig. 3a) and in more natural conditions, with intact females together with food (Supplementary Fig. 3b). Most individual components of the courtship ritual were affected in *Ir84a* mutant flies (Supplementary Fig. 3a). These defects were rescued with a *UAS-Ir84a* transgene, confirming that they result from the absence of IR84a in OSNs (Fig. 3a and Supplementary Fig. 3a, b). The observed reduction in male heterosexual courtship index (~50%) is highly comparable to the phenotype of flies in which all FRU^M-positive OSNs are silenced⁵, suggesting that IR84a-expressing neurons are the major olfactory *fru^M* channel contributing to this behaviour. Residual courtship is presumably stimulated by other sensory modalities, such as taste^{16,17}. Male wild-type *D. melanogaster* also show a low level of courtship towards other males, and we found that this homosexual courtship was also markedly reduced in *Ir84a^{GAL4}* mutants (Fig. 3b). By contrast, *Ir84a^{GAL4}* mutant females did not show overt defects in reproductive behaviours, including copulation latency, success or duration (Fig. 3c).

In innate olfactory preference assays, *Ir84a^{GAL4}* mutant flies still show robust avoidance of acetic acid, indicating that they do not have a general impairment in sensory detection (Supplementary Fig. 3c). By contrast, we did not observe any obvious responses of flies to phenylacetic acid (Supplementary Fig. 3d), suggesting that this food-derived odour is not a volatile stimulus that attracts flies but is a salient cue at close range. Notably, phenylacetic acid has a low vapour pressure (~0.005 mm Hg at 25 °C) compared with other fruit volatiles (for example, ethyl butyrate, which has a vapour pressure of 15 mm Hg at 25 °C). The observation that courtship is reduced in *Ir84a^{GAL4}* mutants in assays in which only small amounts of phenylacetic acid are present on fly cuticles (Supplementary Table 1) raises the possibility that spontaneous activity of these neurons also contributes to establishing a basal courtship level, which is abolished in the absence of IR84a (Fig. 2e, f).

To test whether IR84a ligands are sufficient to promote courtship, we adapted our assay by using killed female objects (which males court at only low levels) and by replacing the base of the chamber with gauze, beneath which we placed a filter paper treated with odour or solvent. Perfuming with phenylacetic acid nearly doubled the courtship index of wild-type flies compared with a solvent control (Fig. 3d). This effect was abolished in *Ir84a^{GAL4}* mutants and could be restored, albeit not fully, by introducing a *UAS-Ir84a* transgene. By contrast, ethyl butyrate, which does not activate IR84a (Supplementary Fig. 2b), did not increase courtship (Fig. 3d). We also perfumed the courtship chamber with *Drosophila* food—which contains phenylacetic acid (Supplementary Table 1)—and observed that this complex olfactory stimulus induced IR84a-dependent increases in male courtship behaviour (Fig. 3e).

The other *fru^M*-expressing OSN populations express either OR67d, which is a receptor for the antiaphrodisiac male pheromone

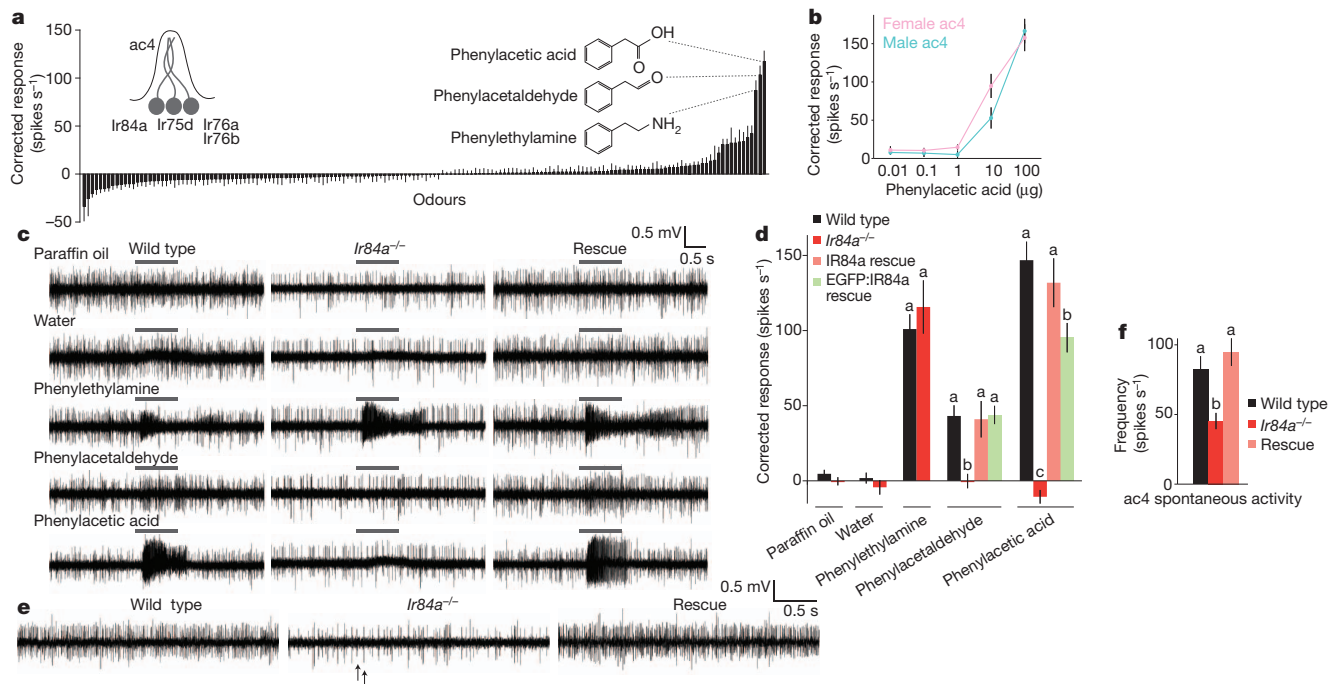


Figure 2 | Essential requirement for IR84a for spontaneous and odour-evoked neuronal responses. **a**, An odour ligand screen in ac4 sensilla. The mean evoked responses of ac4 sensilla neurons to a panel of 163 odours (representing the summed activities of the IR84a-expressing, IR75d-expressing, and IR76a- plus IR76b-expressing neurons). Data are presented as mean \pm s.e.m.; $n \geq 4$, mixed genders. An enlarged version of this histogram with all of the odour names indicated is provided in Supplementary Fig. 2b. The structures of the best three agonists are shown. **b**, The dose response curve to phenylacetic acid of male and female ac4 sensilla. Data are presented as mean \pm s.e.m.; $n = 12$ for each gender. The doses reflect the total mass of phenylacetic acid (diluted in solvent) in the odour syringe. The responses of the male and female sensilla are not significantly different at any concentration (analysis of variance (ANOVA), $P > 0.0571$). A higher concentration of phenylacetic acid could not be prepared because of its limited solubility. **c**, Representative traces of recordings of ac4 sensilla stimulated with the indicated odour stimuli (10 μ l phenylacetic acid (10 μ g μ l⁻¹) or 10 μ l (1% (v/v)) other odours) in wild-type (*w*¹¹¹⁸), *Ir84a*^{-/-} mutant (*Ir84a*^{GAL4}/*Ir84a*^{GAL4}) and *Ir84a* rescue (*UAS-Ir84a*/+; *Ir84a*^{GAL4}/*Ir84a*^{GAL4}) flies. Paraffin oil and water

are solvent controls. The grey bars above the traces mark the stimulus time (1 s). **d**, Quantification of mean responses to the indicated odours for the experiment shown in **c**. Data are presented as mean \pm s.e.m.; $n \geq 11$, mixed genders. The restoration of phenylacetaldehyde and phenylacetic acid responses by the expression of EGFP:IR84a (*UAS-EGFP:Ir84a*/+; *Ir84a*^{GAL4}/*Ir84a*^{GAL4}) is also shown (green). For each odour stimulus, bars labelled with different letters are significantly different from each other (phenylethylamine ANOVA, $P = 0.4861$; phenylacetaldehyde ANOVA, $P = 0.0003$; phenylacetic acid ANOVA, $P < 0.0001$). For phenylacetic acid, EGFP:IR84a rescue is partial. **e**, Representative traces of spontaneous activity in ac4 sensilla in wild-type, *Ir84a*^{-/-} mutant (*Ir84a*^{GAL4}/*Ir84a*^{GAL4}) and *Ir84a* rescue (*UAS-Ir84a*/+; *Ir84a*^{GAL4}/*Ir84a*^{GAL4}) flies. The arrows mark two distinct spike amplitudes discernible in the sparser spike trains in *Ir84a* mutant sensilla; these are likely to correspond to the IR75d- and the IR76a- plus IR76b-expressing neurons. **f**, Quantification of the mean spontaneous activity in ac4 sensilla in the experiment shown in **e**. Data are presented as mean \pm s.e.m.; $n \geq 12$, mixed genders. Bars labelled with different letters are significantly different (ANOVA, $P = 0.0005$).

cis-vaccenyl acetate¹⁸, or OR47b, which is activated by unidentified fly-derived odours from both sexes¹¹ and may participate in mate localization^{19,20}. We examined how IR84a sensory information is integrated with these pheromonal pathways, by visualizing the axons of projection neurons innervating the VL2a (IR84a), VA1lm (OR47b) and DA1 (OR67d) glomeruli, which carry sensory information to the mushroom body and lateral horn²¹ (Fig. 4a). We registered images of single-labelled projection neurons of different glomerular classes onto a common reference brain (see Methods). DA1 and VA1lm excitatory projection neurons target an anterior-ventral pheromone-processing region of the lateral horn²², which is segregated from projection neurons that are responsive to general food odours²². Importantly, we found that VL2a projection neurons—and no other IR-expressing projection neuron class⁹—are highly interdigitated with pheromone pathways and not food pathways (Fig. 4a, b). Indeed, VL2a projection neuron axon terminals overlap more strongly with VA1lm projection neurons than any of the other 44 projection neuron classes (see Methods and Supplementary Fig. 4b, c), consistent with projection neurons of both of these classes transmitting courtship-promoting sensory signals^{19,20}. The VL2a, DA1 and VA1lm inhibitory projection neurons were observed to overlap to a similar extent (Supplementary Fig. 4a–c). The anatomical convergence of combinations of excitatory and inhibitory inputs from VL2a, VA1lm and DA1 projection neurons

may allow the integration of olfactory signals by *fru*^M-expressing third-order neurons^{23,24} to control male courtship behaviour.

Many olfactory IRs are conserved in insects²⁵ and may detect odours that are important for all species. By contrast, although IR84a orthologues are present in ecologically diverse drosophilids, they are absent from other Diptera and more divergent insects (Supplementary Fig. 5a). In the cactophilic species *Drosophila mojavensis*, we identified coeloconic sensilla with neurons that are responsive to phenylacetic acid and phenylacetaldehyde on their anterior antennal surface (similar to ac4 sensilla in *D. melanogaster*) (Supplementary Fig. 5b). Thus, IR84a may have a conserved, drosophilid-specific function.

Despite the widely held assumption of the existence of volatile chemicals that promote courtship in *Drosophila*^{4,5}, behavioural evidence for long-range pheromones is inconclusive^{26,27}, and no female-specific volatile compound that activates male OSNs has been identified¹¹. Our characterization of IR84a identifies an olfactory receptor that is expressed in FRU^M-positive neurons and is required to promote male courtship. Surprisingly, this receptor is not activated by fly-derived odours but rather by aromatic compounds that are present in the vegetal substrates in which fruitflies feed, breed and oviposit⁸. Thus, the IR84a pathway may promote male courtship in the presence of food, complementing the functions of pheromone receptors in regulating mate choice (Supplementary Fig. 5c). This model can account for

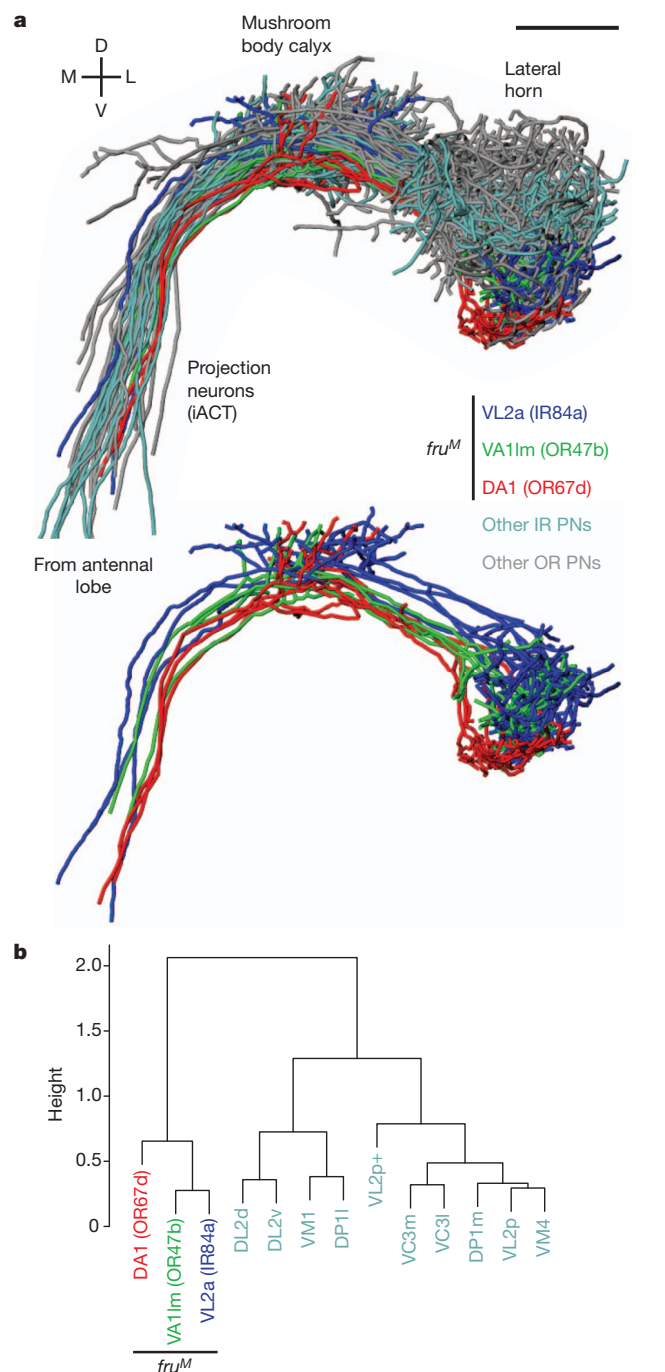
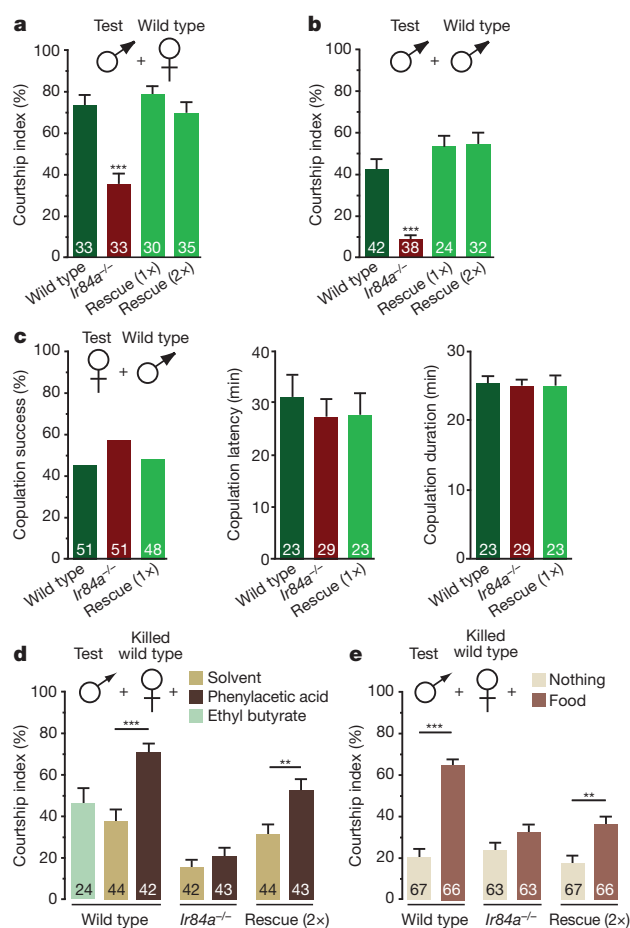


Figure 4 | Anatomical integration of VL2a ($IR84a$ -expressing) projection neurons in the pheromone processing centre. **a, Top, three-dimensional rendering of registered axonal projections of projection neurons (PNs) receiving input from the VL2a glomerulus ($IR84a$ -expressing) (dark blue), the VA1Im glomerulus ($OR47b$ -expressing) (green) and the DA1 glomerulus ($OR67d$ -expressing) (red) ($n = 4$ for each class). Other classes of projection neuron are shown in cyan (nine IRs^9) and grey (27 ORs^{22}) ($n = 1$ for each class). These projection neurons are cholinergic and project to both the mushroom body and the lateral horn through the inner antennocerebral tract (iACT); a small minority of other projection neurons are GABAergic (γ -aminobutyric acid-containing) and project exclusively to the lateral horn through the middle ACT (Supplementary Fig. 4a). Bottom, when the other projection neurons are removed, the extensive co-mingling of VL2a and VA1Im projection neurons is evident. Scale bar, 25 μm . D, dorsal; L, lateral; M, medial; V, ventral. **b**, Cluster analysis of normalized axon overlap scores (y axis) using Ward’s method (see Methods) demonstrates that VL2a projection neurons cluster with projection neurons from other fru^M -expressing glomeruli rather than with other IR -expressing projection neurons. Colours are as in **a**.**

the widespread observations that *D. melanogaster* and other drosophilids mate predominantly on their food substrates^{8,28}. Whereas many insects and other animal classes use long-range sex pheromones to attract potential mates¹, the evolution of IR84a in fruitflies has provided an alternative (although not necessarily exclusive) olfactory mechanism to unite males with females by integrating food-sensing neurons with the circuitry controlling sexual behaviour. Whether other animals have dedicated sensory pathways for environmental 'aphrodisiacs' remains an open question.

METHODS SUMMARY

Genetic and molecular biological manipulations and histological and electrophysiological analyses were performed essentially as described previously^{6,9,10}. Responses to *Drosophila*-derived stimuli were measured by adapting previous protocols¹¹. *Drosophila* extracts and food or fruit extracts were made using CH₂Cl₂ solvent and were analysed by gas chromatography–mass spectrometry on a GC-MS-QP2010 apparatus (Shimadzu). Courtship tests were performed at 25 °C in round observation chambers (1.2 cm in diameter and 0.4 cm deep) under far-red light and were video recorded and annotated manually by observers who were blinded to the genotypes. For perfumed behavioural assays, the courtship chamber was placed on top of an identical chamber, with the chambers separated by muslin gauze to avoid contact between the flies in the top chamber and the chemical stimuli placed in the bottom chamber. Olfactory responses were measured using a Y-maze assay. Projection neurons from several sources (see Methods) were traced, where necessary, using Amira software (<http://www.amira.com>) and registered, by adapting previous approaches²². The degree of overlap between the axon terminals of different projection neurons was measured using the three-dimensional convex hull formed by each set of terminals. Cluster analysis of these data was performed using Ward's method.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 June 2010; accepted 3 August 2011.

Published online 28 September 2011.

- Wyatt, T. D. *Pheromones and Animal Behaviour: Communication by Smell and Taste* (Oxford Univ. Press, 2003).
- Dickson, B. J. Wired for sex: the neurobiology of *Drosophila* mating decisions. *Science* **322**, 904–909 (2008).
- Demir, E. & Dickson, B. J. *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell* **121**, 785–794 (2005).
- Manoli, D. S. *et al.* Male-specific *fruitless* specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* **436**, 395–400 (2005).
- Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L. & Dickson, B. J. Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* **121**, 795–807 (2005).
- Benton, R., Vannice, K. S., Gomez-Diaz, C. & Vosshall, L. B. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* **136**, 149–162 (2009).
- Wightman, F. & Lighty, D. L. Identification of phenylacetic acid as a natural auxin in the shoots of higher plants. *Physiol. Plant.* **55**, 17–24 (1982).
- Markow, T. A. & O'Grady, P. Reproductive ecology of *Drosophila*. *Funct. Ecol.* **22**, 747–759 (2008).
- Silbering, A. F. *et al.* Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *J. Neurosci.* doi:10.1523/JNEUROSCI.2360-11.2011 (21 September 2011).
- Abuin, L. *et al.* Functional architecture of olfactory ionotropic glutamate receptors. *Neuron* **69**, 44–60 (2011).
- van der Goes van Naters, W. & Carlson, J. R. Receptors and neurons for fly odors in *Drosophila*. *Curr. Biol.* **17**, 606–612 (2007).
- Ferveur, J. F. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behav. Genet.* **35**, 279–295 (2005).
- Yao, C. A., Ignell, R. & Carlson, J. R. Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *J. Neurosci.* **25**, 8359–8367 (2005).

- Barata, A. *et al.* Analytical and sensorial characterization of the aroma of wines produced with sour rotten grapes using GC-O and GC-MS: identification of key aroma compounds. *J. Agric. Food Chem.* **59**, 2543–2553 (2011).
- Kim, J., Jeon, C. O. & Park, W. A green fluorescent protein-based whole-cell bioreporter for the detection of phenylacetic acid. *J. Microbiol. Biotechnol.* **17**, 1727–1732 (2007).
- Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019–1029 (2003).
- Kohatsu, S., Koganezawa, M. & Yamamoto, D. Female contact activates male-specific interneurons that trigger stereotypic courtship behavior in *Drosophila*. *Neuron* **69**, 498–508 (2011).
- Kurtovic, A., Widmer, A. & Dickson, B. J. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* **446**, 542–546 (2007).
- Root, C. M. *et al.* A presynaptic gain control mechanism fine-tunes olfactory behavior. *Neuron* **59**, 311–321 (2008).
- Wang, L. *et al.* Hierarchical chemosensory regulation of male–male social interactions in *Drosophila*. *Nature Neurosci.* **14**, 757–762 (2011).
- Vosshall, L. B. & Stocker, R. F. Molecular architecture of smell and taste in *Drosophila*. *Annu. Rev. Neurosci.* **30**, 505–533 (2007).
- Jefferis, G. S. *et al.* Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* **128**, 1187–1203 (2007).
- Cachero, S., Ostrovsky, A. D., Yu, J. Y., Dickson, B. J. & Jefferis, G. S. Sexual dimorphism in the fly brain. *Curr. Biol.* **20**, 1589–1601 (2010).
- Yu, J. Y., Kanai, M. I., Demir, E., Jefferis, G. S. & Dickson, B. J. Cellular organization of the neural circuit that drives *Drosophila* courtship behavior. *Curr. Biol.* **20**, 1602–1614 (2010).
- Croset, V. *et al.* Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **6**, e1001064 (2010).
- Ewing, A. W. & Manning, A. The effect of exogenous scent on the mating of *Drosophila melanogaster*. *Anim. Behav.* **11**, 596–598 (1963).
- Shorey, H. H. & Bartell, R. J. Role of a volatile female sex pheromone in stimulating male courtship behaviour in *Drosophila melanogaster*. *Anim. Behav.* **18**, 159–164 (1970).
- Spieth, H. T. Courtship behavior in *Drosophila*. *Annu. Rev. Entomol.* **19**, 385–405 (1974).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to B. Dickson, A. Hofbauer, T. Lee, the Bloomington *Drosophila* Stock Center, the *Drosophila* Species Stock Center and the Developmental Studies Hybridoma Bank for provision of plasmid vectors, *Drosophila* strains and antibodies. We are also grateful to A. Wong, J. Wang, R. Axel, H.-H. Yu and T. Lee for sharing raw image data. We thank D. Featherstone, J.-F. Ferveur, T. Kaweck, L. Keller, S. Martin and members of the Benton laboratory for comments on the manuscript. Y.G., J.-P.F. and J.C. are supported by the Centre National de la Recherche Scientifique (CNRS), the Agence Nationale de la Recherche (ANR; JCJC, GGCB-2010) and the Conseil Régional de Bourgogne (FABER). R.R. was supported by a Roche Research Foundation fellowship. G.S.X.E.J. is supported by the Medical Research Council and a European Research Council Starting Investigator Grant. Research in R.B.'s laboratory is supported by the University of Lausanne, a European Research Council Starting Independent Researcher Grant and the Swiss National Science Foundation.

Author Contributions Y.G. and R.B. conceived the project. Y.G. performed the gene-targeting screen, contributed to the histological analysis and performed most of the behavioural experiments. R.R. performed the electrophysiological odour response screen, the fly odour stimulation assays and the phylogenetic analyses. J.-P.F. performed the chemical analysis. L.A. assisted in the generation and characterization of transgenic flies and contributed to the histological analysis. J.C. contributed to the behavioural experiments. G.S.X.E.J. performed the analysis of projection neurons. R.B. generated the DNA constructs, performed all other electrophysiological analyses and wrote the paper with contributions from Y.G., R.R., J.-P.F. and G.S.X.E.J.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.B. (Richard.Benton@unil.ch).

METHODS

Drosophila strains. *Drosophila* stocks were maintained on a standard corn flour, yeast and agar medium under a 12 h light and 12 h dark cycle at 25 °C. We used the following mutant and transgenic strains: *Or35a-GAL4* (ref. 13), *Ir84a-GAL4* (ref. 9), *UAS-Ir84a^Δ*, *UAS-EGFP-Ir84a^Δ*, *UAS-mCD8:GFP²⁹*, *fru^{GAL4}* (ref. 5), *fruP1-GAL4* (ref. 4), *lexAop-rCD2::GFP*, *UAS-mCD8* (ref. 30), *UAS-DTP³¹*, *70FLP, 70I-SceI/Cyo*, *70FLP* and *70I-CreP³²*. For behavioural analysis, all flies carrying mutant and transgenic chromosomes were backcrossed to an isogenized *w¹¹¹⁸* control line for five generations. The sequenced wild-type *D. melanogaster* strain (Bloomington *Drosophila* Stock Center number 2057) was used for the ac4 odour screen, and Canton-S was used as the courtship object in behavioural experiments. Wild-type *D. mojavensis* was obtained from the *Drosophila* Species Stock Center (University of California, San Diego; stock number 15081-1352.22).

Molecular biology. New transgenes were constructed as follows, using standard methods. For the *Ir84a* *GAL4* targeting construct, 5,087 base pairs (bp) genomic DNA immediately upstream of the *Ir84a* start codon was amplified and subcloned as 1,924-bp and 3,163-bp fragments either side of the I-SceI site upstream of *GAL4* in the pED36 targeting vector³. Similarly, a 964-bp genomic fragment immediately downstream of the *Ir84a* stop codon was cloned downstream of *GAL4*. For the *Ir84a* promoter-*lexA* construct, the same 1,964 bp *Ir84a* promoter fragment used in the *Ir84a-GAL4* transgene⁹ was cloned upstream of *lexA:VP16-SV40* (ref. 30) in the vector pattB³³.

Gene targeting. Targeted replacement of *Ir84a* by *GAL4* was performed essentially as described previously¹⁸. A transgenic line containing the *Ir84a* *GAL4* targeting construct on the X chromosome was used as a donor for targeting crosses in which heat-shock-inducible transgenic FLP and I-SceI enzymes were used to excise and linearize the targeting fragment. Reinsertion was detected by the presence of a FLP-insensitive *white⁺* marker. From ~193,000 flies screened, five targeted insertions were recovered and verified by PCR as duplicates of the *Ir84a* locus. One duplicate was resolved with I-CreI, selecting for loss of *white⁺*. From 218 *white⁻* flies, 1 was found to contain the desired precise *Ir84a* to *GAL4* replacement.

Histology. Immunofluorescence on antennal cryosections was performed as described previously^{10,34}. Immunofluorescence on whole mount brains, and combined fluorescent RNA *in situ* hybridization and immunofluorescence on antennal sections, was performed as described previously⁹. The primary antibodies used were mouse monoclonal nc82 (1:10 dilution; provided by A. Hofbauer), mouse monoclonal 21A6 (1:200; Developmental Studies Hybridoma Bank), rabbit anti-GFP (1:1,000; Invitrogen), mouse anti-GFP (1:500; Invitrogen), rat anti-mCD8 (1:50; Caltag), anti-DIG-POD (1:500; Roche). The secondary antibodies used were Alexa Fluor 488-, Cy3- and Alexa Fluor 647-conjugated goat anti-mouse, anti-rabbit or anti-rat IgG (Molecular Probes and Jackson ImmunoResearch). These were diluted to 1:100 and 1:1,000 for whole mount brains and antennal sections, respectively. All microscopy was performed using an LSM 510 laser scanning confocal microscope (Zeiss).

Electrophysiology. Extracellular recordings in single sensilla of 1–8-day-old flies were performed and quantified as described previously⁶. Odours were diluted to 1% (v/v) unless otherwise noted. The CAS numbers, solvents, sources and purities for the odorants used are provided in Supplementary Table 2. Corrected responses were quantified by counting all spikes in ac4 sensilla in a 0.5 s window from 150–200 ms after the stimulus trigger. This delay period was defined precisely for each recording session with a control odour response and mainly represents the time for the odour to reach the preparation. From this spike value, we subtracted the number of spontaneous spikes in a 0.5 s window before stimulation and doubled the result to obtain spikes s⁻¹. Spontaneous activity was quantified by counting the spikes in a 5 s window without stimulus and then dividing by five to obtain spikes s⁻¹. After verifying that the responses were normally distributed, we compared all genotypes for a given experiment by analysis of variance (ANOVA), with the genotype as the main effect, and adjusted the alpha level for planned post hoc means comparisons.

To assess the responses of live flies to volatile compounds, ~100–120 male or virgin female flies were placed in a 5 ml glass pipette (witeg Labortechnik). The pipette was then closed at each end with cotton gauze, and an air pulse (3 s) was blown through it. As controls, an empty pipette was used or a pipette loaded with 10 µl phenylacetic acid (0.1 µg µl⁻¹ in paraffin oil, as 1 µl drops on ten individual 2 mm² filter papers). The responses were quantified by counting all of the spikes in a 1 s window from 500 ms after the odour stimulus trigger and then subtracting the number of spontaneous spikes in a 1 s window before stimulation. To assess the responses to *D. melanogaster* cuticular hydrocarbons, we prepared extracts in 50 µl CH₂Cl₂ as described below (except without the *n*-C13 standard) and then evaporated the solvent under N₂ and diluted the extract in paraffin oil to a concentration of 10 fly equivalents µl⁻¹. These extracts (1 µl) (or control stimuli) were placed on the tip of a borosilicate glass capillary (World Precision Instruments) and brought

into close proximity (~5 mm) to the antenna. The changes in spiking frequency were quantified by measuring the spontaneous spiking frequency in spikes s⁻¹ in a 3 s window before the stimulus was brought close to the antenna and then subtracting this value from the spiking frequency in a 3 s window when the stimulus was placed close to the antenna. This second 3 s window typically started 3–5 s after stimulus placement, a time frame when the responses to phenylacetic acid were most robust.

Extract preparation and chemical analysis. As phenylacetaldehyde and phenylacetic acid are polar compounds, we used CH₂Cl₂ for all extractions. *D. melanogaster* extracts were prepared by incubating five to ten 5-day-old flies for 5 min in vials containing 50 µl CH₂Cl₂ and 100 ng *n*-C13 (as an internal standard) at room temperature. After the flies were removed, the extracts were stored at -20 °C until analysis. For standard fly food and fruit extracts, 2 g material was crushed in 20 ml vials containing 4 ml CH₂Cl₂ and 200 ng *n*-C13 as an internal standard. After vortexing for 10 min at room temperature, the extracts were filtered through glass wool and concentrated to ~100 µl under a gentle flow of N₂. The extracts were stored at -20 °C until analysis.

Because of the difficulty of detecting phenylacetic acid and phenylacetaldehyde at low concentration in the extracts, we tested three column types with different polarities: a CP-Sil 5 CB (apolar type, 25 m × 0.25 mm internal diameter, 0.12 µm film thickness; Varian), a CP-Wax 58 CB (polar type, 25 m × 0.25 mm internal diameter, 0.20 µm film thickness; Varian) and a VF-1ms (medium polarity type, 20 m × 0.15 mm internal diameter, 0.15 µm film thickness; Varian). Only the last two columns gave satisfactory quantitative results, and these were used for all subsequent analyses. The extracts were analysed using a GC-MS-QP2010 apparatus (Shimadzu) in splitless mode, fitted with a VF-1ms (*D. melanogaster* extracts) or a CP-Wax 58 CB (standard fly food, banana and prickly-pear extracts) fused silica capillary column. The columns were held isothermally at 40 °C for 2 min, then the temperature was programmed to increase at a rate of 3 °C min⁻¹ to 240 °C (CP-Wax 58CB) or 300 °C (VF-1ms). Helium was used as the carrier gas at a linear velocity of 47 cm s⁻¹. The injector port was set at 280 °C. The mass spectrometer was operated at 70 eV, and scanning was performed from 29 to 600 AMU at 0.5 scans s⁻¹. The injection split was opened 1 min after the injection. Phenylacetic acid and phenylacetaldehyde were identified using their retention time and their fragmentation patterns; diagnostic ions were compared with both the NIST/EPA/NIH Mass Spectral Library and the mass spectrum of synthetic chemical standards (Sigma-Aldrich) analysed under the same conditions. For quantitative analyses of phenylacetic acid and phenylacetaldehyde, the response factor was determined for each of the compounds and *n*-C13 at 0.1, 0.5, 1, 5 and 10 ng on both types of column. For analysis of extracts from flies grown on minimal medium, late second or young third instar larvae reared on standard food medium were washed six times in distilled water to remove any traces of medium. Larvae were then transferred to a minimal growth medium (0.1 M sucrose and 1% agarose). At emergence, adults were separated by sex and aged in groups of 20 flies on minimal growth medium for 5 days, before extraction as described above.

Courtship assays. Male and virgin female flies were collected at eclosion. They were kept in vials, individually for subjects or in groups of 10 for objects, for 5–9 days before behavioural assays. The vials contained fresh food and were kept under a 12 h light and 12 h dark cycle at 25 °C. Behavioural assays were performed at approximately the same time each morning. Single-pair courtship tests were performed in round observation chambers (1.2 cm in diameter and 0.4 cm deep, with nine-chamber plexiglass courtship wheels) under far-red light (to eliminate the contribution of visual cues) at 25 °C. For most experiments, object females (or males) were decapitated with a clean razor blade (to avoid any reciprocal interaction during courtship) and were left to 'recover' for ~10 min before introduction into the courtship chamber. We also tested males for their courtship behaviour in more natural conditions: using intact live virgin Canton-S females, together with ~90 mg fly food medium (corresponding to approximately one-quarter of the courtship chamber volume). Male courtship behaviour towards these different types of object was recorded for 10 min with an HDR-SR10 digital video camera (Sony). After recording, videos were analysed by a researcher who was blinded to the genotypes, using Annotation software (SaySoSoft). The courtship index was calculated as the percentage of time that a male courted the object during a 10-min period. Female copulation behaviour was quantified by direct observation for 1 h without video recording; copulation latency represents the time to the beginning of copulation from the moment that the virgin female and the male are placed in the chamber; copulation success is the percentage of flies that copulate; and copulation duration is the total time of copulation from the start to the end of copulation.

For perfumed assays, the courtship chamber was placed on top of an identical chamber, and the chambers were separated by muslin gauze to avoid direct contact between the flies in the top chamber and the chemical stimuli in the bottom chamber (food, 50 ± 2 mg standard fly food; solvent, 10 µl paraffin oil (on a filter paper or aluminium foil pad); phenylacetic acid, 10 µl ~35 µg µl⁻¹ solution in

paraffin oil (that is, ~ 2.5 μmol total odour); ethyl butyrate, 10 μl 1% (v/v) solution in paraffin oil (that is, ~ 0.8 μmol total odour)). Single object males were placed in the upper chamber with a killed intact virgin wild-type female at 25 °C. The female had been anaesthetized with CO₂, placed in an Eppendorf tube and killed by submerging the tube in liquid nitrogen. The female was then warmed for ~ 10 min at 25 °C in a Petri dish before being placed in the courtship chamber. Males court killed objects only at low levels, facilitating the determination of the enhancing effect of perfuming.

Olfactory assays. Male olfactory responses were tested using an adapted Y-maze assay³⁵; in brief, three-way plastic tube connectors (Reactolab) were used to join three glass vials by way of 1 ml pipette tips that pierced the cotton vial tops, forming a tightly sealed Y maze. The narrow ends of the pipette tips were cut and oriented in each vial top to form one 'loading' vial and two 'trap' vials, which contained 40 μl odour (10% (v/v) acetic acid in water (that is, ~ 70 μmol total odour) or ~ 35 μg μl^{-1} phenylacetic acid in paraffin oil (that is, ~ 10 μmol total odour) or 40 μl corresponding solvent on filter paper. Ten 5–9-day-old males (starved for 18–21 h in glass tubes with water on filter paper) were introduced without CO₂ anaesthesia into the loading vial and were allowed 2 h at 25 °C to choose to enter the trap vial containing the odour or the solvent. The resultant olfactory index was obtained with the following formula: (number in the odour tube – number in the solvent tube)/total number of loaded males.

Projection neuron analysis. Projection neuron images (275) from four different sources were brought into a common reference space for analysis^{22,36–38}. The common template (Cell07) was from ref. 22, so 236 projection neuron images from that study did not require further processing. The anti-Discs large channel of 34 confocal images from ref. 38 was registered to a single female template brain (FlyCircuit ID 6475, FC6475), which was chosen for good quality staining and imaging, using the CMTK toolkit (<http://www.nitrc.org/projects/cmtk>) as described previously²³. The FC6475 template brain was then registered to the Cell07 brain using landmarks registration based on 24 manually chosen anatomical landmarks visible in both anti-Discs large and nc82 staining. This landmarks registration used the pnreg command of the IRTK toolkit (<http://www.doc.ic.ac.uk/~dr/software>), which uses nonlinear third-order B-spline registration³⁹. The fiducial registration error, measured as the root mean squared deviation, was 5.1 μm after affine registration and 2.4 μm after warping registration, corresponding to per axis accuracies of 2.9 and 1.4 μm , respectively. The final per axis registration accuracy for independent landmarks (not used during the registration) was 3.3 μm , which is comparable to that of previous studies^{22,24}.

Three VL2a projection neuron images from ref. 37 were registered to the Cell07 template by choosing 12–14 landmarks in the nc82 channel present in each stack. Landmarks registration used the pnreg command of IRTK. Two VL2a projection neuron images from ref. 36 were registered to the Cell07 template using 10–14 landmarks. Although those brains had weak nc82 staining, the only reliable marker was GH146-GAL4-driven expression of a CD2 reporter. We therefore used a two-channel confocal image of an nc82-stained brain with GH146-driven mCD8:GFP expression that had been registered by way of its nc82 channel to the Cell07 template to help choose landmarks. Landmarks registration again used the pnreg command of IRTK. The fiducial registration error (root mean squared deviation) for the five brains varied between 3.6 and 4.9 μm (affine) and 1.8 and 3.1 μm (warp).

Neuronal tracing was carried out in Amira (<http://www.amira.com>) with the hxskeletonize plug-in⁴⁰. The three-dimensional coordinates of neuronal tracings were then transformed from their respective original image coordinates using the CMTK gregxform tool (FlyCircuit images \rightarrow FC6475 template) and the IRTK ptransformation tool (FC6475 \rightarrow Cell07, three VL2a images \rightarrow FC6475). The underlying tools were called using custom code written in R (<http://www.r-project.org>), which was also used to analyse the transformed neuronal tracings (using the AnalysisSuite R package)²². The final visualization was carried out in Amira using Tcl scripts to load, show/hide and colour neuron tracings and to take snapshots. R and Tcl source code, landmark files, template brain images and neuronal tracings will be available on publication at the Jefferis laboratory website (<http://flybrain.mrc-lmb.cam.ac.uk>).

Analysis of axonal overlap. The degree of overlap between the axon terminals of different projection neurons was measured by making use of the three-dimensional convex hull of each terminal (using the Qhull library⁴¹ as exposed by the R geometry package). It may help to note that for a set of two-dimensional points in the plane, the convex hull can be obtained by stretching an elastic band to encompass the whole object and then releasing it. To obtain an overlap score for a projection

neuron pair (A,B), we calculated three convex hulls, $H(A)$, $H(B)$ and $H(A,B)$, where $H(A,B)$ is the convex hull for all of the points in A and B. We then calculated a normalized overlap score as $s(A,B) = (H(A) + H(B) - H(A,B))/(H(A) + H(B))$.

If A and B are identical (that is, complete overlap), the score will be 0.5. Less overlap will result in lower scores, with negative scores when the amount of intervening space between the axon terminals exceeds the overlap (if any). We first calculated an overlap score for all 275 neurons in the data set. We then aggregated the scores for all pairwise combinations of the 46 projection neuron classes in the data set and calculated the median score for each combination. This resulted in a symmetrical distance matrix with $46 \times 45/2 = 1,035$ unique off-diagonal entries. The median overlap score for VL2a and VA11m projection neurons was 0.224, which was at the 97.7th centile: that is, the overlap between these two classes was among the strongest 2.3% in the data set. VL2a was the neuronal class with the strongest overlap with VA11m (and vice versa). Cluster analysis of these data was performed with Ward's method, using the hclust function of R. Overlap scores, s , were converted to a distance, d , suitable for clustering by the simple transform $d = 0.5 - s$.

The VL2a projection neuron image data were obtained in separate studies from the bulk of the projection neuron data, so it is natural to ask whether this could have had some effect on our analysis. However, the consistent results for VL2a projection neuron images obtained from two different laboratories argue very strongly against this possibility. Visual inspection indicated that there was excellent overlap between the VL2a neurons from the two sources, and the pairwise overlap scores between and within the two groups were almost identical (median 0.284 and median 0.301, corresponding to the 98.0th and 98.8th centiles in Supplementary Fig. 4b). As an additional check, we repeated the axon overlap analysis and clustering with affine registered VL2a neurons (affine registration is more robust, although it is less accurate in the ideal situation). This had a negligible effect on the VL2a/VA11m median overlap score (0.197 and 0.224, corresponding to the 96.2th centile and 97.7th centile, respectively) and had no effect on the clustering results (G.S.X.E.J., data not shown).

Phylogenetic analysis. The phylogenetic tree of selected antennal-expressed IR orthologous groups²⁵ was constructed using PhyML⁴² and is based on the most conserved columns of an amino acid alignment generated with MUSCLE⁴³. The phylogeny was rooted using IR25a. iTOL⁴⁴ was used to view and graphically edit the tree.

29. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
30. Lai, S. L. & Lee, T. Genetic mosaic with dual binary transcriptional systems in *Drosophila*. *Nature Neurosci.* **9**, 703–709 (2006).
31. Han, D. D., Stein, D. & Stevens, L. M. Investigating the function of follicular subpopulations during *Drosophila* oogenesis through hormone-dependent enhancer-targeted cell ablation. *Development* **127**, 573–583 (2000).
32. Rong, Y. S. et al. Targeted mutagenesis by homologous recombination in *D. melanogaster*. *Genes Dev.* **16**, 1568–1581 (2002).
33. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific ϕ C31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
34. Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol.* **4**, e20 (2006).
35. Alcorta, E. & Rubio, J. Intrapopulation variation of olfactory responses in *Drosophila melanogaster*. *Behav. Genet.* **19**, 285–299 (1989).
36. Wong, A. M., Wang, J. W. & Axel, R. Spatial representation of the glomerular map in the *Drosophila* protocerebrum. *Cell* **109**, 229–241 (2002).
37. Yu, H. H. et al. A complete developmental sequence of a *Drosophila* neuronal lineage as revealed by twin-spot MARCM. *PLoS Biol.* **8**, e1000461 (2010).
38. Chiang, A. S. et al. Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution. *Curr. Biol.* **21**, 1–11 (2011).
39. Rueckert, D. et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* **18**, 712–721 (1999).
40. Evers, J. F., Schmitt, S., Sibila, M. & Duch, C. Progress in functional neuroanatomy: precise automatic geometric reconstruction of neuronal morphology from confocal image stacks. *J. Neurophysiol.* **93**, 2331–2342 (2005).
41. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The Quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
42. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
43. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
44. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).

Molecular organization of vomeronasal chemoreception

Yoh Isogai^{1,2}, Sheng Si¹, Lorena Pont-Lezica^{1†}, Taralyn Tan¹, Vikrant Kapoor¹, Venkatesh N. Murthy¹ & Catherine Dulac^{1,2}

The vomeronasal organ (VNO) has a key role in mediating the social and defensive responses of many terrestrial vertebrates to species- and sex-specific chemosignals¹. More than 250 putative pheromone receptors have been identified in the mouse VNO^{2,3}, but the nature of the signals detected by individual VNO receptors has not yet been elucidated. To gain insight into the molecular logic of VNO detection leading to mating, aggression or defensive responses, we sought to uncover the response profiles of individual vomeronasal receptors to a wide range of animal cues. Here we describe the repertoire of behaviourally and physiologically relevant stimuli detected by a large number of individual vomeronasal receptors in mice, and define a global map of vomeronasal signal detection. We demonstrate that the two classes (V1R and V2R) of vomeronasal receptors use fundamentally different strategies to encode chemosensory information, and that distinct receptor subfamilies have evolved towards the specific recognition of certain animal groups or chemical structures. The association of large subsets of vomeronasal receptors with cognate, ethologically and physiologically relevant stimuli establishes the molecular foundation of vomeronasal information coding, and opens new avenues for further investigating the neural mechanisms underlying behaviour specificity.

The discovery of large receptor families mediating olfactory and vomeronasal chemosensation has offered a unique opportunity to decode the molecular logic by which environmental information influences animal behaviour^{3,4}. The VNO of rodents has a critical role in identifying sex- and species-specific chemical cues and in mediating mating, territorial aggression, defensive responses to predators and associated endocrine changes^{1,5}. With rare exceptions^{6–8}, the molecular identity of VNO receptors (VRs) recognizing distinct animal cues is unknown, thus limiting the ability to explore the sensory mechanisms underlying behavioural specificity. Prior studies suggested that vomeronasal detection is extremely sensitive and narrowly tuned to male, female or heterospecific cues^{5,9–11}, but they have not allowed the identification of the activated receptors. We describe here a robust and high-throughput molecular readout of vomeronasal activation that enabled us to uncover the receptor specificity of 88 individual VRs to a vast range of ethologically relevant cues. These results establish the molecular and functional framework underlying vomeronasal signalling.

In initial experiments, we exposed female mice to clean bedding and to bedding used by male mice, and assessed the upregulation of the immediate early genes (IEGs) *Arc*, *c-Fos*, *c-Jun*, *Egr1*, *FosB* and *Nr4a1* by *in situ* hybridization on VNO tissue. Our data show that the sensitivity of *Egr1* induction following exposure to chemosignals far exceeds that of other IEGs (Fig. 1a, b) (60.1 ± 7.1 cells per 0.2 mm^2 for *Egr1*, 7.9 ± 1.9 cells per 0.2 mm^2 for *c-Fos*). Indeed *c-Fos*, an IEG used in previous VNO stimulation studies, labels only a subset of *Egr1*-positive cells (Supplementary Fig. 1). In *TrpC2*^{−/−} mutants, in which VNO activation is genetically impaired¹², *Egr1* induction after semiochemical exposure is completely abolished ($n = 3$), demonstrating the specificity of *Egr1* activation following sensory stimulation (Fig. 1c).

We then exposed animals to 29 distinct ethologically relevant cues^{5,13}. Male and female bedding from different mouse subspecies and wild-derived strains, as well as a variety of heterospecific cues from sympatric competitors and predators, robustly induced *Egr1* expression in the VNO (Fig. 2a). Remarkably, food-related insect stimuli and cues from presumably neutral species such as woodchuck failed to generate VNO activation.

V1R and V2R neurons were equally activated by a large variety of stimuli, as judged by co-labelling of *Egr1* with *Gα_{i2}*, a marker of V1R-expressing neurons^{14,15} (Fig. 2b, Supplementary Fig. 2a). Interestingly, simultaneous exposure to multiple cues from the same class of animals (for example, *Peromyscus* species, reptiles, or predatory birds) did not significantly increase the number of *Egr1*-positive cells when compared to activation by a single stimulus ($P > 0.4$, two-tailed *t*-test when the strongest of each stimulus class was compared to the corresponding mix), indicating that neuronal populations activated by related animals are largely overlapping (Fig. 2a). In contrast, simultaneous exposure to all heterospecific stimuli significantly increased *Egr1*-positive cells from 5% to 10% per cue to up to ~30% ($P < 0.01$, two-tailed *t*-test), indicating that distinct heterospecific cues have different response profiles. Moreover, whereas mouse bedding activated 5% to 7% of VNO neurons in animals of the opposite sex, mixes of conspecific and heterospecific scents activated ~35% of neurons (Fig. 2a), suggesting that receptors activated by both types of cues are also largely distinct.

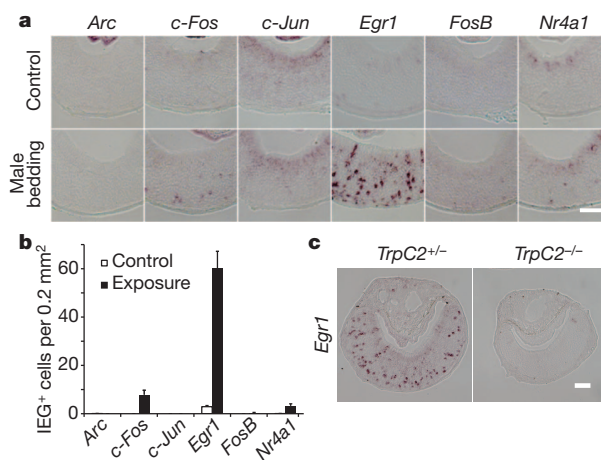
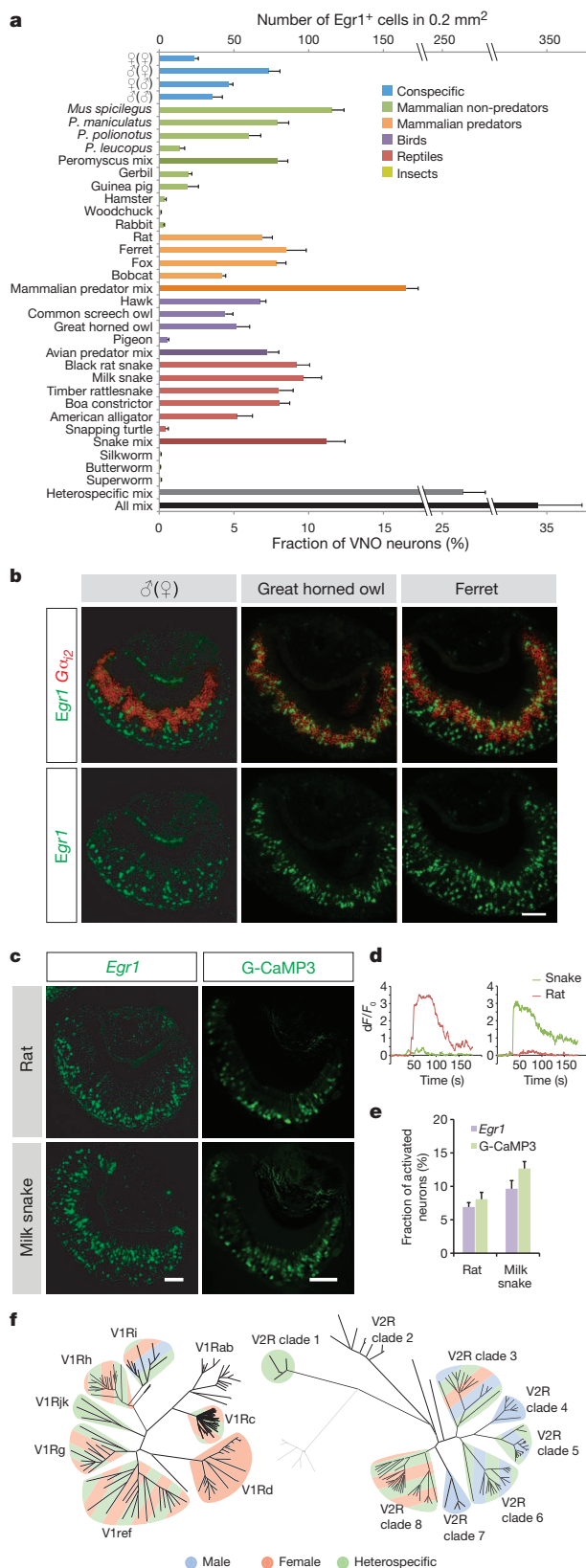


Figure 1 | *Egr1* expression is robustly induced by pheromone-evoked VNO neuronal activation. Female CD-1 mice were exposed to clean or male mouse bedding and their VNOs analysed for expression of various immediate early genes (IEGs). **a**, *In situ* hybridization with RNA probes to *Arc*, *c-Fos*, *c-Jun*, *Egr1*, *FosB* and *Nr4a1*. **b**, Numbers of IEG-positive cells after bedding exposure (10 sections per VNO, $n = 3$ animals). Error bars, s.e.m. **c**, *TrpC2*, a cation channel involved in VNO signal transduction is required for *Egr1* induction. Female *TrpC2*^{+/−} or *TrpC2*^{−/−} mice were exposed to male conspecific bedding and *Egr1* expression was visualized in the VNO. Scale bar, 100 μm .

¹Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. ²Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. [†]Present address: Ecole Normale Supérieure, Paris 75005, France.



To assess *Egr1* as a readout of VNO activation, we compared it to cue-evoked neuronal responses visualized by the genetically encoded calcium indicator, G-CaMP3 (ref. 16). Strikingly, *Egr1* and G-CaMP3 reported remarkably similar patterns of activities in the basal, or basal-plus-apical VNO neuroepithelium following exposure to rat and snake stimuli, respectively (Fig. 2c–e), confirming *Egr1* induction as an exquisitely sensitive and accurate marker of VNO neuronal activation.

Figure 2 | Widespread activation of VNO receptors by conspecific and heterospecific cues. **a**, Survey of ethologically relevant vomeronasal stimuli. Vomeronasal neural activation on exposure to conspecific and heterospecific cues was visualized by *Egr1* induction and quantified. Detection of female cues by males is designated as (ϕ/σ) . Unless specified, female mice were used. Mixed heterospecific cues activated *Egr1* in significantly more cells than individual stimuli ($P < 0.01$, two-tailed *t*-test). Co-exposure to heterospecific and conspecific stimuli (all mix, $n = 6$) resulted in significantly more *Egr1*-positive cells ($P < 0.05$, two-tailed *t*-test). **b**, Widespread activation of VNO neurons by conspecific and heterospecific cues. Shown are *in situ* hybridization results with probes against *Gx12* (red) and *Egr1* (green). **c**, Comparison between *Egr1* and G-CaMP3-evoked signal in response to rat or milk snake chemosignals. G-CaMP3 images are 10-s averages of ΔF frames within stimulus period. **d**, Differential response profiles of neurons to rat or snake signals. Stimuli were perfused from 20 s to 60 s. **e**, Quantitative comparison between *Egr1* and G-CaMP3-evoked signals. The percentage of activated cells identified by G-CaMP3 ($n = 356$ cells for rat stimuli, $n = 566$ cells for snake stimuli, 9 VNO slices from 3 animals) among those responsive to 40 mM KCl was plotted in the graph. Data for *Egr1* were taken from **a**. The difference between *Egr1* and G-CaMP3 was not statistically significant ($P > 0.1$, two-tailed *t*-test). **f**, Clade-level maps of V1R (left) and V2R (right) activation show distinct clade specificity for male, female or heterospecific cues. Hatched patterns indicate response to multiple types of cues. Error bars, s.e.m. Scale bars, 100 μm .

Next, we developed a high-throughput platform to uncover the receptors activated by specific cues. With the exception of widely expressed V2R2 receptors¹⁷, vomeronasal neurons are thought to express a unique receptor gene from the V1Rs or V2Rs. We generated 209 RNA probes that specifically identify individual or subgroups of VRs by *in situ* hybridization. A collection of clade-specific probes was designed to target all receptor sequences within each of the eight distinct V1R or V2R clades (Fig. 2f). Probes with higher specificity that readily distinguish a single or few closely related VR sequences were designed on the basis of divergent 5'-UTR/intron¹⁸ and 3'-UTR regions in VR genes. The specificity of these probes for closely related VRs was validated by dual colour *in situ* hybridization (Supplementary Fig. 3). Although detecting all VRs at single gene resolution was technically impossible, our probes targeted 139 distinct VRs with a specificity of a single (or at most a few) gene.

We then used a hierarchical approach to systematically uncover VRs activated by distinct cues (Supplementary Figs 2b, 4). First, the co-expression of *Egr1* with either $G\alpha_{i2}$, $G\alpha_o$ or formyl peptide receptors (FPRs)^{19,20} identified the nature of the activated neurons as expressing a V1R, V2R or FPR, respectively. Most stimuli activated both V1R- and V2R-expressing neurons, while a few activated only V1R- (hawk and owls) or V2R-expressing cells (rat, fox and male mouse cues in females) (Supplementary Table 1). We found no activation of FPR-expressing cells. We then assessed the specific V1R or V2R clades associated with the activated neurons (Fig. 2f, Supplementary Fig. 2c). Interestingly, some clades appeared specifically stimulated by a distinct class of cues, for example, V1Rd and V2R clades 4 and 7 by sex-specific cues. Subsequently, receptor specific probes were used to unmask the exact molecular identity of the *Egr1*-positive cells. By collecting data from 9,948 VNO slices, each containing approximately 1,000 neurons, we succeeded in the identification of 88 receptors (56 V1Rs and 32 V2Rs, 78 single and 10 unresolved receptors) associated with distinct cues (Supplementary Fig. 5, Supplementary Table 1, 2). Importantly, these receptors span most V1R and V2R clades, thus generating the most comprehensive functional map of vomeronasal receptors to date.

The vomeronasal system plays an essential part in regulating sex-specific behaviours. Previous reports suggest that vomeronasal neurons detect sex-specific cues in mouse urine, tear and saliva^{9,10,13,21,22}, and Vmn2r116 (or V2Rp5) was identified as detecting the male pheromone ESP1 (ref. 6; Supplementary Fig. 6). Our strategy uncovered 28 receptors (25 single, 3 unresolved) detecting mouse cues, among which 26 detecting sex-specific cues (Fig. 3a–c, Supplementary Table 1). Only two receptors (V1ri9, V1ri10) responded to both male and female mouse cues, consistent with the desensitization of IEG induction *in*

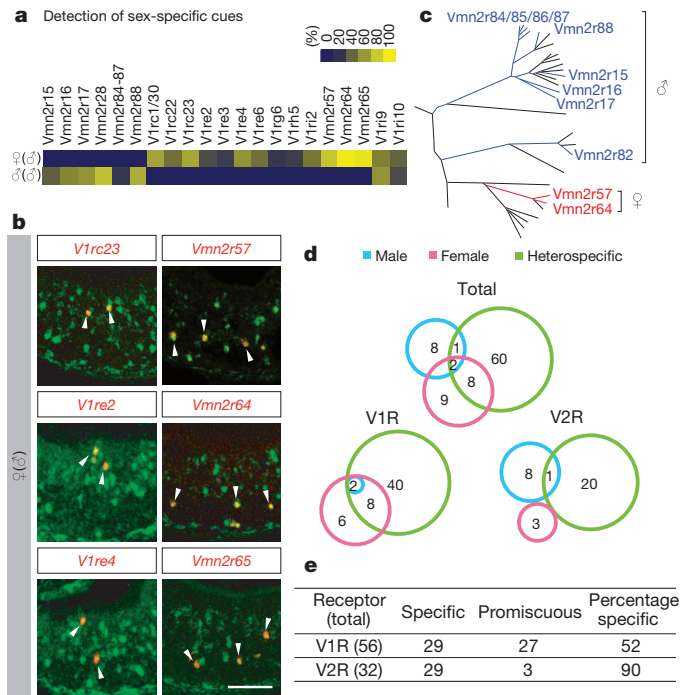


Figure 3 | Receptor responses to sex-specific cues. **a**, **b**, Male and female mouse cues are each detected by a specific subset of V1Rs and V2Rs. **a**, Heat maps representing the co-localization between *Egr1* and representative vomeronasal receptor genes (yellow, 100% overlap; dark blue, 0% overlap). **b**, *In situ* hybridization of *Egr1* (green) and individual receptors (red), with arrows marking co-localization of *Egr1* and receptor signals. Scale bars, 100 μ m. **c**, Clade organization of V2Rs detecting male (blue) or female (red) cues. **d**, Receptors detecting male (blue), female (red) and heterospecific (green) cues are largely distinct. **e**, V1Rs and V2Rs display distinct specificity. Shown are the numbers of receptors that detect unique types of cues (specific) versus multiple types (promiscuous) among the following categories: male, female, mammalian non-predator, mammalian predator, reptile, and avian predator.

vivo by self-secreted stimuli⁶. Four receptors (V1re2, V1re3, V1re6, V1rg6) were selectively activated by female cues in males and females, while a larger set of V1Rs and V2Rs responded to female cues only in males (Fig. 3a–c, Supplementary Table 1). In addition, responses to male-specific signals involved Vmn2r116, Vmn2r28, Vmn2r15, Vmn2r16 and Vmn2r17 in males and females, Vmn2r66 and Vmn2r82 in females, and Vmn2r84/85/86/87 and Vmn2r88 in males (Fig. 3a–c, Supplementary Table 1). Interestingly, no V1R was found to specifically respond to male cues. Thus, consistent with a previous report⁹, the detection of sex-specific cues appears to rely on a small and specific subset of VNO neurons, the identity of which is now clearly established. This molecular logic is likely to underlie the initiation of sex-dependent behavioural interactions, such as male–male aggression and mating behaviours.

Vomeronasal detection of heterospecific cues, or kairomones, is involved in adaptive defensive behaviours^{5,13,23}. Indeed, rat bedding induces robust avoidance to the predator cues in *TrpC2*^{+/−} but not in *TrpC2*^{−/−} animals (Fig. 4g, h). Moreover, *TrpC2*^{−/−} animals exhibited abnormal ingestive behaviour of the predator bedding, suggesting that VNO inputs also inhibit foraging^{24,25} (Supplementary Fig. 7).

We report here the identity of 71 (63 single, 8 unresolved) receptors activated by heterospecific scents. Consistent with the distinct behavioural outputs generated by pheromones and kairomones, we found that only 11 receptors were common to both types of cues, whereas 60 were uniquely activated by heterospecific stimuli, and 17 by mouse cues only (Fig. 3d). The detection of kairomones thus emerges as a major function of the VNO^{5,13}. The identity of one of the identified receptor population for the detection of predator cues was confirmed independently by *Egr1* activation in cells expressing YFP under the V1Rh7

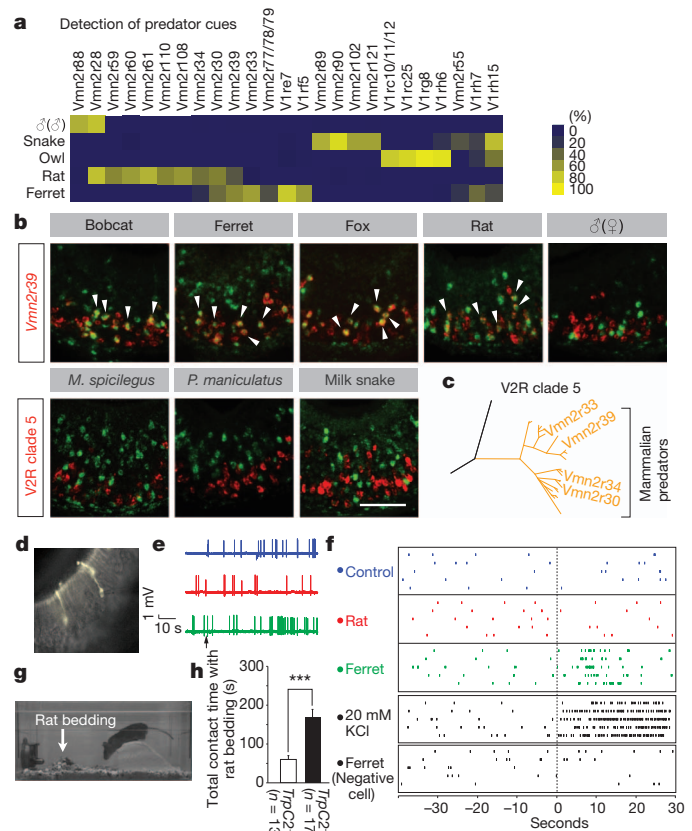


Figure 4 | Receptor responses to heterospecific cues. **a**, **b**, Predator cues are detected by a specific subset of V1Rs and V2Rs. **a**, Heat map representing the co-localization between *Egr1* and representative vomeronasal receptor genes (colour coding as Fig. 3a). **b**, *In situ* hybridization of *Egr1* (green) and vomeronasal receptors (red), with arrows marking co-localization of *Egr1* and receptor signals. Scale bar, 100 μ m. **c**, Mammalian predator cues commonly activate V2R clade 5 receptors. Owing to high homology among V2R clade 5 genes, the Vmn2r30, 33, 34, 39 probes detect multiple receptors. **d**, Fluorescence image showing a patched V1rh7-YFP neuron. **e**, Loose-patch recordings of a V1rh7-YFP neuron. The arrow indicates perfusion start. **f**, Spike raster for three different VNO neurons, showing responses of a V1rh7-YFP neuron to ferret, but not to rat stimuli, and no response of a V1rh7-YFP-negative neuron to ferret stimuli. The stimulus perfusion started at −30 s and lasted 20 s. **g**, **h**, Rat bedding (arrow) elicits robust avoidance behaviours in control *TrpC2*^{+/−} mice, but significantly less in *TrpC2*^{−/−} mice lacking VNO activity. ****P* < 0.0001 (two-tailed Student's *t*-test). Error bars, s.e.m. (*TrpC2*^{+/−}, *n* = 13; *TrpC2*^{−/−}, *n* = 17).

promoter²⁶ (Supplementary Fig. 8). Further, loose patch recording of V1Rh7-YFP expressing neurons demonstrated significant increase in firing rates following exposure to ferret, but not to rat stimuli (1.732 ± 0.170 Hz for ferret, 0.420 ± 0.061 Hz for rat, *n* = 4) (Fig. 4d–f, Supplementary Fig. 9).

Remarkably, some receptors show unique association with distinct classes of predators. Vmn2r89 and Vmn2r121 were exclusively activated by scents from snakes, V1rc10/11/12 by owls. Also, up to 70% of V2R clade 5 neurons were activated by every mammalian predator tested, but not by sympatric non-predators (Fig. 4a–c, Supplementary Fig. 5, 10). Moreover, each predator cue generated a distinct receptor signature: for example, rat stimuli activate Vmn2r59, Vmn2r60, Vmn2r61, Vmn2r108 and Vmn2r110, all within clade 8, whereas ferret cues activate V1rf5 and Vmn2r77/78/79, suggesting that the mouse VNO has the sensory machinery to discriminate predator species.

We then searched for receptors detecting the sympatric species *Mus spicilegus* and *Mus musculus*, which diverged evolutionarily ~1.5 million years ago and do not breed in the wild^{27,28}. Receptors activated by *M. spicilegus* and *M. musculus* male cues appear mostly distinct, though often closely related (Supplementary Figs 5, 11). For

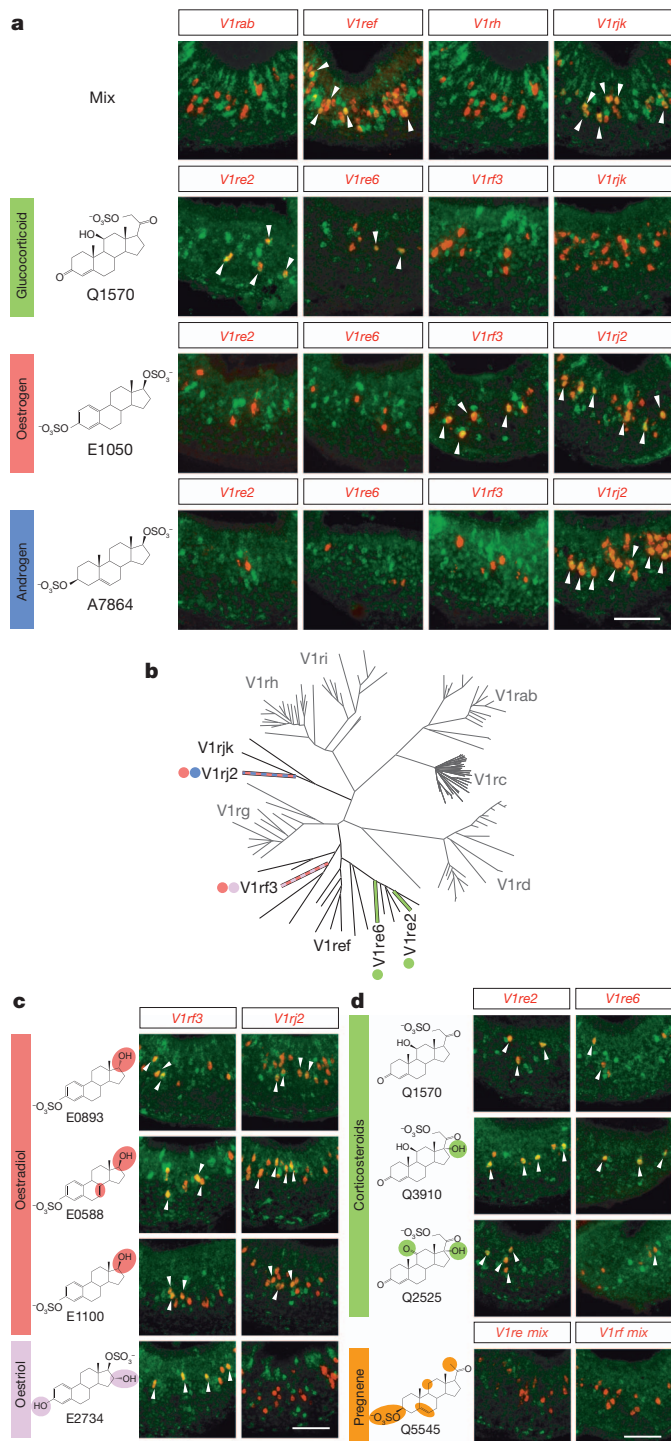


Figure 5 | Sulphated steroids detection by V1Rs. **a**, V1Ref and V1Rjk clade-specific probes (red) co-localize with *Egr1* (green) after VNO stimulation by a mix of steroids containing glucocorticoids such as Q1570 (green), oestrogens such as E1050 (red) and androgens such as A7864 (blue). Each of these compounds on its own elicits activity in distinct populations of vomeronasal neurons (V1re2, V1re6, V1rf3 and V1rj2), also represented in the molecular tree of V1R receptors (**b**). Specific receptors detecting each steroid are indicated by dots, using the same colour scheme as **a** and **c**. **c**, The three distinct oestradiols (red) activate both V1rf3 and V1rj2 whereas the oestriol (purple) only activates V1rf3. **d**, The sulphate group position in pregnenes (corticosteroids in green, pregnenolone in orange) determines the specificity of ligand detection by V1re2 and V1re6. Differences in chemical structures among tested compounds are highlighted by coloured circles. Arrowheads mark co-localization between *Egr1* and receptor signals. Scale bar, 100 μ m.

example, Vmn2r8/9 and Vmn2r11, activated by *M. spicilegus*, and Vmn2r15, Vmn2r16 and Vmn2r17, activated by *M. musculus*, belong to clade 6 (Supplementary Fig. 11b). Likewise, Vmn2r69 (activated by *M. spicilegus*) and Vmn2r66 (activated by *M. musculus*) belong to clade 3. Thus, through the activation of specialized receptors, *M. musculus* may readily discriminate scents emitted by closely related but reproductively incompatible species, a property that could be linked to the reproductive isolation of these species.

V1Rs and V2Rs are associated with segregated neural pathways²⁹, raising the possibility that fundamental functional differences may exist between the two families. Remarkably, our data suggest that V1Rs and V2Rs display different receptor properties. Nearly half of the V1Rs (27 out of 56) exhibit generalized activation by multiple cues (Fig. 3e), including signals with apparent conflicting behavioural significance. For example, receptors within the V1Rh, V1Rc and V1Re clades were activated by mouse, predator and non-predator cues (Supplementary Tables 1 and 2, Supplementary Fig. 12). In contrast, most V2Rs (29 out of 32) are activated by cues reflecting a unique ethological context such as a male, a female, or a given type of predator or non-predator. In addition, hierarchical clustering across all identified receptors revealed clear segregation between V1Rs and V2Rs (Supplementary Fig. 5). These results suggest that V1R and V2R pathways may encode different types of information: individual V2Rs appear uniquely poised to encode information about the identity of emitters with clear behavioural significance—for example, the sex of a conspecific or the nature (predator or competitor) of a heterospecific. In contrast, individual V1Rs may encode other forms of biologically relevant information.

To gain further insight into the molecular logic of V1R-mediated signalling, we investigated the detection of sulphated steroids, which are thought to account for 80% of VNO neuronal activation by female urine³⁰ (probably through V1Rs¹¹). Our data show that, when male mice were exposed to a mix of synthetic steroid sulphates, receptors from V1Ref and V1Rjk clades were specifically activated (Fig. 5a, b). We then tested individual compounds to attempt the pairing of specific steroid ligands with their cognate receptors. Corticosterone-21 sulphate (Q1570), a compound in female urine³⁰, strongly activated V1re2 and more weakly V1re6 cells (Fig. 5a, b). Both receptors were shown in earlier experiments to be specifically activated by female cues (Fig. 3a). In addition, we uncovered strong activation of V1rf3 by 17 β -oestradiol sulphate (E1050) and V1rj2 by both E1050 and 5-androstene-3 β , 17 β -diol disulphate (A7864) (Fig. 5a), although these two receptors were not activated by female bedding, indicating that these steroids are not secreted under normal conditions.

Thus, our approach efficiently achieved single compound resolution, offering the unique opportunity to test the receptor specificity to a variety of individual chemicals. We further tested four sulphated oestrogen compounds structurally related to E1050, and three additional sulphated pregnenes structurally related to Q1570. V1rf3 appeared broadly selective to oestradiols, oestriols and related stereoisomers, regardless of sulphate positions, but did not respond to androgens or glucocorticoids (Fig. 5c). Interestingly, no other V1rf receptor was activated by these ligands. In contrast, V1rj2 was activated by androgens and oestradiols but not oestriols. Similarly, V1re2 and V1re6 selectively detected corticosteroids (Fig. 5d). Therefore, V1R receptors can distinguish distinct structural classes of steroids. Androgens, oestrogens and glucocorticoids are ubiquitous though sensitive reporters of the animal endocrine state. Our results thus suggest that V1Rs may serve as detectors of the physiological status of an animal.

In conclusion, our data have begun to uncover the molecular logic by which vomeronasal receptors of different families, clades and receptor sequences extract biological information and trigger appropriate behavioural responses to animals of a given sex, species and physiological status. The collection of receptors uncovered in this study provides a molecular foundation to further dissect the neural circuits governing social and sexual communication in rodents.

METHODS SUMMARY

Stimulus exposure was conducted by introducing a subject animal (male or female CD-1 mice, 8 to 14 weeks old) in a fresh cage containing distinct animal cues for 30 min (for Fig. 1) or 40 min (for Figs 2–5). The dissected VNOs were embedded in OCT (Tissue-Tek) and frozen in dry ice. Cryosections (16 µm) of VNO were subjected to RNA *in situ* hybridization using IEG and VR probes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 January; accepted 8 August 2011.

Published online 21 September 2011.

- Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4**, 551–562 (2003).
- Zhang, X., Marcucci, F. & Firestein, S. High-throughput microarray detection of vomeronasal receptor gene expression in rodents. *Front. Neuroscience* **4**, 164 (2010).
- Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141**, 692–703 (2010).
- Haga, S. *et al.* The male mouse pheromone ESP1 enhances female sexual receptive behaviour through a specific vomeronasal receptor. *Nature* **466**, 118–122 (2010).
- Leinders-Zufall, T., Ishii, T., Mombaerts, P., Zufall, F. & Boehm, T. Structural requirements for the activation of vomeronasal sensory neurons by MHC peptides. *Nature Neurosci.* **12**, 1551–1558 (2009).
- Boschat, C. *et al.* Pheromone detection mediated by a V1r vomeronasal receptor. *Nature Neurosci.* **5**, 1261–1262 (2002).
- He, J., Ma, L., Kim, S., Nakai, J. & Yu, C. R. Encoding gender and individual information in the mouse vomeronasal organ. *Science* **320**, 535–538 (2008).
- Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
- Holekamp, T. F., Turaga, D. & Holy, T. E. Fast three-dimensional fluorescence imaging of activity in neural populations by objective-coupled planar illumination microscopy. *Neuron* **57**, 661–672 (2008).
- Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
- Ben-Shaul, Y., Katz, L. C., Mooney, R. & Dulac, C. *In vivo* vomeronasal stimulation reveals sensory encoding of conspecific and allospecific cues by the mouse accessory olfactory bulb. *Proc. Natl Acad. Sci. USA* **107**, 5172–5177 (2010).
- Berghard, A. & Buck, L. B. Sensory transduction in vomeronasal neurons: evidence for G α_o , G α_{i2} , and adenylyl cyclase II as major components of a pheromone signaling cascade. *J. Neurosci.* **16**, 909–918 (1996).
- Jia, C. & Halpern, M. Subclasses of vomeronasal receptor neurons: differential expression of G proteins (G α_{i2} and G α_o) and segregated projections to the accessory olfactory bulb. *Brain Res.* **719**, 117–128 (1996).
- Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
- Martini, S., Silvotti, L., Shirazi, A., Ryba, N. J. & Tirindelli, R. Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ. *J. Neurosci.* **21**, 843–848 (2001).
- Stewart, R. & Lane, R. P. V1R promoters are well conserved and exhibit common putative regulatory motifs. *BMC Genomics* **8**, 253 (2007).
- Liberles, S. D. *et al.* Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proc. Natl Acad. Sci. USA* **106**, 9842–9847 (2009).
- Rivière, S., Challet, L., Flügge, D., Spehr, M. & Rodriguez, I. Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **459**, 574–577 (2009).
- Holy, T. E., Dulac, C. & Meister, M. Responses of vomeronasal neurons to natural stimuli. *Science* **289**, 1569–1572 (2000).
- Taha, M., McMillon, R., Napier, A. & Wekesa, K. S. Extracts from salivary glands stimulate aggression and inositol-1, 4, 5-triphosphate (IP3) production in the vomeronasal organ of mice. *Physiol. Behav.* **98**, 147–155 (2009).
- Samuelson, C. L. & Meredith, M. The vomeronasal organ is required for the male mouse medial amygdala response to chemical-communication signals, as assessed by immediate early gene expression. *Neuroscience* **164**, 1468–1476 (2009).
- Brown, J., Kotler, B., Smith, R. & Wirtz, W. The effects of owl predation on the foraging behavior of heteromyid rodents. *Oecologia* **76**, 408–415 (1988).
- Sundell, J. *et al.* Variation in predation risk and vole feeding behaviour: a field test of the risk allocation hypothesis. *Oecologia* **139**, 157–162 (2004).
- Wagner, S., Gresser, A. L., Torello, A. T. & Dulac, C. A multireceptor genetic approach uncovers an ordered integration of VNO sensory inputs in the accessory olfactory bulb. *Neuron* **50**, 697–709 (2006).
- Chevret, P., Veyrunes, F. & Britton-Davidian, J. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biol. J. Linn. Soc.* **84**, 417–427 (2005).
- Guénet, J. L. & Bonhomme, F. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* **19**, 24–31 (2003).
- Dulac, C. & Wagner, S. Genetic analysis of brain circuits underlying pheromone signaling. *Annu. Rev. Genet.* **40**, 449–467 (2006).
- Nodari, F. *et al.* Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. *J. Neurosci.* **28**, 6407–6418 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge H. Fisher, H. Hoekstra, E. Kay, M. Kirchgessner, N. Uchida, A. Wang, X.-D. Wang, B. Watson, W. Tong, Harvard Museum of Natural History, Harvard Concord Field Station, Museum of Science, Boston, and New England Wildlife Center, for providing stimulus materials used in this study, L. Looger for the G-CaMP3 construct, M. Wienisch, F. Markopoulos and D. Mak for help with electrophysiology and imaging experiments, and B. Goetze and the Harvard Center for Biological Imaging for help with microscopy. We also thank members of the Dulac laboratory for critical reading of the manuscript, S. Andreeva for technical support and R. Hellmiss for help with figure artwork. This work was supported by the NIDCD at the National Institute of Health, the Howard Hughes Medical Institute and the Damon Runyon Cancer Research Foundation (Y.I., DRG-1981-08).

Author Contributions Y.I. and C.D. designed the study. Y.I., S.S. and T.T. designed and generated RNA probes, performed RNA *in situ* hybridization, and analysed data. L.P.-L. performed pilot experiments for data shown in Fig. 1 and produced recombinant ESP1. Y.I. and V.K. performed calcium imaging and electrophysiology. V.N.M. supervised physiology experiments. Y.I. and C.D. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.D. (dulac@fas.harvard.edu).

METHODS

Sampling of animal stimuli. Bedding materials used in this study were all freshly sampled from cages that house live animals (Harvard University, Harvard Museum of Natural History, Harvard Concord Field Station, Tufts University, Museum of Science, Boston, and New England Wildlife Center). Soiled bedding represents the most complete stimulus source for animals, and is also of ecological relevance. Bedding materials typically absorb a wide range of chemical stimuli excreted by animals, such as urine, faeces, saliva, fur, and other gland secretions. Since different animals are housed in different environments, we flexibly adjusted the sampling procedures. For instance, chemosignals emitted by heterospecific mammals and birds (male rat, female fox, male ferret, female bobcat, male *Peromyscus*, male *M. spicilegus*, male and female gerbils, male and female hamsters, male and female rabbits, woodchuck, pigeon, red tailed hawk, screech owl, and great horned owl) were sampled as soiled bedding (paper, woodchips or corn cob). For reptiles, we sampled faeces, urate and other gland secretions absorbed in woodchips or paper. These bedding materials were directly used for exposure experiments (as described separately below). For aquatic animals such as alligators, only faecal pellets were sampled. For insect larvae, live animals were directly used for exposure experiments. Some predators such as snake and predatory birds were fed mice as part of their diet, and we took great care to avoid potential odour contamination. For example, on bedding sampling we avoided areas where mouse carcass was present in animal cages. Second, to sample milk snake odour, which we extensively used for our study, we changed bedding after the feeding to avoid potential odour contamination from mice. We also tested materials from multiple individuals whenever possible. Judging from the number of *Egr1*-positive cells, we did not find extensive individual variability in these samples. If multiple individuals were not available, especially for bobcat, hawk and great horned owl, we tested stimulus samples from different collection dates. We stored these bedding materials at 4 °C for the short term (one week) and at -20 °C for the long term. These materials, even after long term storage at -20 °C when the amount of volatiles was significantly reduced, did not appreciably lose their ability to robustly stimulate vomeronasal neurons.

For conspecific stimuli, to represent a potential diversity of chemical cues emitted by different subspecies of mice, we collected bedding samples from 5 different strains of mice: BALB/c (Jackson Labs), PWD/PhJ (Jackson Labs), CAST/Eij (Jackson Labs), Idaho³¹ and Chuuk³¹, and exposed these samples as a mixture. It is known that mice secrete different vomeronasal cues reflecting their physiological states, for example, different phases of oestrous⁹, prompting us to sample materials freshly from cages that house multiple animals over 1 week. Thus, conspecific stimuli used in our study probably contain chemosignals secreted over different phases of the oestrous cycles. We stored these materials at 4 °C for the short term and -20 °C for the long term.

Stimulus exposure. For most exposure experiments involving bedding stimuli, approximately 50 ml (in volume) of bedding containing animal cues were placed in a clean cage. We introduced a subject mouse (male or female CD-1, from 8 weeks to 14 weeks old, Charles River), which voluntarily made extensive direct contacts with introduced stimuli in freely behaving conditions. The animals were exposed to stimuli for 30 min (for Fig. 1) or 40 min (for Figs 2–5). Subsequently, the dissected VNOs were embedded in OCT (Tissue-Tek) and frozen in dry ice. VNO cryosections (16 µm) were used for RNA *in situ* hybridization using IEG and vomeronasal receptor probes. Control experiments were conducted using fresh bedding in an identical manner. For insect larvae exposure, 3–4 insect larvae were directly introduced to the cages. For alligator stimuli, a few faecal pellets were used. For heterospecific mix exposure experiments, ~100 ml mixture of the following bedding sample was used: *Peromyscus* (*P. maniculatus*, *P. leucopus*, *P. polionotus*), mammalian predators (bobcat, fox, ferret, rat), avian predators (screech owl, great horned owl, red tail hawk), reptiles (rat snake, milk snake, rattlesnake, boa, alligator), and *M. spicilegus*. For pure chemicals such as ESP1 and sulphated steroids, ~5 µl of Ringer's (in mM; 115 NaCl, 5 KCl, 2 CaCl₂, 2 MgCl₂, 25 NaHCO₃ and 5 HEPES) containing the stimuli were directly spotted on each nostril. Recombinant ESP1 was purified as a GST fusion protein overexpressed in *Escherichia coli* using pET41 vector (Novagen), followed by thrombin cleavage to release the ESP1 peptide. 2 µg of the peptide was exposed to each animal.

Sulphated steroid exposure. Steroids were purchased from Steraloids. A mix of steroids (A6940, epitestosterone sodium sulphate; A7864, 5-androsten-3β, 17β-diol disulphate; E1050, 17β-oestradiol sulphate; E0893, 17α-oestradiol sulphate; P3817, allopregnanolone sulphate; P8200, epipregnanolone sulphate, Q1570, corticosterone 21-sulphate; Q3470, deoxycorticosterone 21-glucoside; each at 250 µM in Ringer's) were used for initial screens. Subsequently, individual steroids (Q1570; E1050; A7864; E0893; E0588, 17β-dihydroequilin 3-sodium sulphate; E1100, 17β-oestradiol 3-sulphate; E2734, oestriol 17-sulphate; Q3910, hydrocortisone 21-sodium sulphate; Q2525, cortisone 21-sulphate; Q5545, 3β-hydroxy-5-pregnen-20-one 3-sulphate) were used at 500 µM in Ringer's. 5 µl of steroid

solution were spotted on each nostril of male CD-1 animals (8–14 weeks), and the animals were exposed to steroids for 40 min. Experiments were conducted for at least three animals.

Preparation of RNA probes. For immediate early gene probes, we have cloned complementary DNA of *Arc*, *c-Fos*, *c-Jun*, *Egr1*, *FosB* and *Nr4a1* in approximately 900-base-pair (bp) segments to pCRII-TOPO or pCR4-TOPO vector (Invitrogen). Antisense cRNA probes were synthesized using T3, T7 or Sp6 polymerases (Promega) and digoxigenin (DIG) or fluorescein (FITC) labelling mix (Roche) from PCR templates. All IEG probes consisted of a cocktail of 2–3 probes to cover nearly the full length of these messenger RNAs.

For V1R clade-specific probes, we cloned full length coding sequence of V1R receptors (V1rab: *V1ra1*, *a2*, *a3*, *a4*, *a5*, *a6*, *a7*, *a8*, *b1*, *b2*, *b7*, *b8*, *b9*; V1rc: *V1rc3*, *c8*, *c10*, *c16*, *c28*; V1rd: *V1rd6*, *d9*, *d11*, *d12*, *d14*, *d22*, *Vmn1r167*; V1ref: *V1re1*, *e2*, *e3*, *e4*, *e6*, *e7*, *e8*, *e9*, *e10*, *e11*, *e12*, *e13*, *Vmn1r224*, *f1*, *f2*, *f3*, *f4*, *f5*; V1rh: *V1rh1*, *h20*; V1ri: *V1ri1*, *i3*, *i4*, *i5*, *i6*, *i8*; V1rjk: *V1rj2*, *j3*, *k1*) and combined these probes to generate a clade-specific probe set. For V1rg receptors, ~1 kilobase (kb) 5'-UTR/intron sequences of the following genes were used: *V1rg1*, *g2*, *g3*, *g4*, *g5*, *g6*, *g7*, *g8*, *g9*, *g10*, *g11*, *g12*, *Vmn1r77*, which were combined with *V1rl* cDNA probe to generate the V1Rgl clade probe set.

To generate clade-specific V2R probes, we cloned the first ~900 bp of annotated V2R receptor coding sequence (V2R clade 1: *Vmn2r55*; V2R clade 2: *Vmn2r19*, *Vmn2r20*, *Vmn2r24*; V2R clade 3: *Vmn2r65*, *Vmn2r69*, *Vmn2r76*, *Vmn2r77*; V2R clade 4: *Vmn2r115*; V2R clade 5: *Vmn2r28*, *Vmn2r48*; V2R clade 6: *Vmn2r8*, *Vmn2r15*, *Vmn2r17*, *Vmn2r84*, *Vmn2r89*, *Vmn2r118*; V2R clade 7: *Vmn2r18*, *Vmn2r81*, *Vmn2r83*, *Vmn2r120*; V2R clade 8: *Vmn2r57* 3'-UTR probe, *Vmn2r58*, *Vmn2r63*, *Vmn2r58*, *Vmn2r90*, *Vmn2r93*, *Vmn2r96*, *Vmn2r97*, *Vmn2r99*, *Vmn2r102*, *Vmn2r104*, *Vmn2r105*, *Vmn2r106*, *Vmn2r108*, *Vmn2r110*, and *Vmn2r64* 3'-UTR probe) and combined these probes to generate clade-specific probe sets. To generate cRNA probes specific to individual V1R genes, we cloned ~1 kb 5'-UTR intron sequence of V1R genes to pCRII vector (Invitrogen). To produce cRNA probes specific to individual V2R receptors, we cloned ~600 bp of V2R 3'-UTR segments. These RNA probes were first used to test mRNA expression. We found that some annotated vomeronasal receptor genes did not appear to be expressed, since these RNA probes gave no discernible signals. For all vomeronasal receptor genes, for which we could confirm the expression, we tested the specificity of these probes by dual colour *in situ* hybridization using DIG and FITC probes and used for receptor mapping experiments. Probes generated in our study to detect specific receptors are listed in Supplementary Table 1. The VR nomenclature was based on that of GenBank and Mouse Genome Informatics.

RNA *in situ* hybridization. Single colour RNA *in situ* hybridization was conducted essentially as described³². We used DIG labelled cRNA probes at 2 ng µl⁻¹ and used a hybridization temperature of 65 °C for experiments shown in Fig. 1. For *Egr1* *in situ* hybridization experiments shown in Fig. 2, we used 68 °C as the hybridization temperature. Dual colour fluorescence *in situ* hybridization was conducted in the following steps. First, the tissue was fixed in 4% formaldehyde/1× PBS for 10 min, and washed 3 times with 1× PBS for 3 min each. The tissues were treated with acetylation solution (0.1 M triethanolamine with 2.5 µl ml⁻¹ acetic anhydride) for 10 min. After 3 washes with 1× PBS, each for 5 min, the slide was incubated with the pre-hybridization solution (50% formamide, 5× SSC, 5× Denhardt's, 2.5 mg ml⁻¹ yeast RNA, 0.5 mg ml⁻¹ herring sperm DNA) for 2 h. The hybridization buffer (4% dextran sulphate, Millipore, added to pre-hybridization buffer) containing FITC labelled *Egr1* probes (a cocktail of three probes, each at 50 pg µl⁻¹) and DIG labelled VR probes (at 0.5 ng µl⁻¹ for cDNA probes, and 1 ng µl⁻¹ for 5'-UTR-intron and 3'-UTR probes) was heated at 95 °C for 3 min and immediately chilled on ice for 5 min. The hybridization solution was applied to the slides, which were covered with parafilm and incubated in a sealed chamber for 16 h at 68 °C. Following hybridization, the slides were washed with 5× SSC once for 5 min, and with 0.2× SSC three times, each for 20 min at 68 °C. Slides were washed at room temperature with 0.2× SSC for 5 min and subsequently with TNT buffer (100 mM Tris, pH 7.5, 150 mM NaCl, 0.05% Tween 20) for 5 min.

After the post-hybridization washes, 200 µl of anti-FITC-POD (Roche, at 1/250 dilution in TNB blocking buffer, Perkin-Elmer) was applied and incubated for 3 h at room temperature. Slides were washed with TNT buffer for a total of 1 h, with buffer exchanges every 10 min. The signal was developed using the TSA biotin plus kit (Perkin Elmer), as per manufacturer's protocol. The slides were washed with TNT buffer 3 times, each for 5 min, and subsequently treated with 3% H₂O₂/1× PBS to kill residual peroxidase activity. Slides were washed again 3 times with 1× PBS and TNT, each for 5 min. DIG antibody solution (anti-DIG-POD, Roche, at 1/500 dilution, and Streptavidin-Alexa488, Invitrogen, at 1/250 dilution in TNB buffer) were applied to the slides and incubated overnight at 4 °C. After washing slides with TNT (6 times, 10 min each), the signal was developed using the TSA

Cy3 plus kit (Perkin Elmer) as per manufacturer's protocol. Slides were washed with TNT (3 times, 5 min each and once for 1 h), and tissues were mounted with Vectashield (Vector labs) containing $8 \mu\text{g ml}^{-1}$ DAPI. All the microscopy images were acquired using LSM510 or AxioImager Z2 (Zeiss).

Analysis of *in situ* hybridization images. For single colour *in situ* hybridization images, quantitation was conducted using a minimum of 10 VNO sections per animal and 3 animals (data in Fig. 1) or 3–4 animals (data in Fig. 2). Since we found that 0.2 mm^2 represents areas occupied by medial cryostat sections of the VNO and contain approximately 1,000 VNO cells, we used the average number of *Egr1*-positive cells per 0.2 mm^2 in Fig. 1, and we converted these numbers to percentage of activated neurons among total VNO neurons in Fig. 2. For dual colour *in situ* hybridization, we quantitated the co-localization of *Egr1* and receptor signals over four sections per VNO, for a minimum of three animals. We then calculated the percentage of activated neurons in specific receptor neurons, for each odour class, and generated a co-localization matrix. In many cases, we found that individual receptor mapping is unnecessary when the hierarchical screen can unequivocally demonstrate that there are no activated neurons in specific receptor clades. In these cases, we input zero values to the co-localization matrix. For hierarchical clustering of the co-localization matrix, we used the Cluster program (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>), with average linkage in Euclidian distance. To generate the clustering diagram in Supplementary Fig. 4, we calculated the average number of receptor neurons per receptor in 12 sections and used this as a weight. The heat map and clustering dendrogram were generated using the Java Treeview program (<http://jtreeview.sourceforge.net/>).

Behavioural assay. Male *TrpC2* mice (+/– or –/–, 8–14 weeks old, ref. 12) were single-housed three days before the experiment in a manner blind to the experimenter. The behaviour experiment was conducted by introducing 50 ml volume of fresh or rat bedding to one side of the cage, away from the nest area. The behaviours of the subject mice were video recorded and total contact time as well as ingestive behaviour were scored by an individual blind to the genotype. We defined ingestive behaviour as animals engaged in ingestion while holding a food pellet with two forepaws.

Generation of OMP-GCaMP3 transgenic line. pJOMP plasmid containing the rat olfactory marker protein (OMP) genomic sequence³³ was modified so that the GCaMP3 ORF sequence completely replaces the OMP ORF. Linearized vector was used for pronuclear injection (performed by Harvard Genome Modification Facility), and transgenic founders were further backcrossed to C57Bl/6 mice to establish an OMP-GCaMP3 line. This line expresses the transgene uniformly throughout the vomeronasal epithelium and showed no sign of reported cell toxicity¹⁵.

Calcium imaging on VNO slices. Calcium imaging was carried out essentially as described⁹, using 5–8-week-old female OMP-GCaMP3 mice. The VNOs were acutely dissected, separated from bones, and embedded in 4% low melting point agar in mACSF (in mM; 130 NaCl, 5 KCl, 1 MgCl₂, 2.5 CaCl₂, 1.25 NaH₂PO₄, 25

NaHCO₃, 10 glucose). The coronal vibratome sections (200 μm) were cut, and slices were kept in continuously oxygenated mACSF for up to 8 h at 25 °C. The flow rate of the stimulus was approximately 0.3 ml min^{-1} , and we delivered stimulus for 40 s. All imaging was conducted at 25 °C. The fluorescence changes due to calcium transients were monitored using a LSM710 microscope with a GaAsP detector (Zeiss). We used a 1:100 dilution of freshly sampled rat urine from 2–6-month-old CD male rats (Charles River) in mACSF. For snake stimuli, shredded snake bedding (that is, paper) was extracted with mACSF, filtered and used for stimulation. To quantify the number of activated cells, we first generated ΔF images by subtracting an average of 20 s frames corresponding to initial resting period from the raw images. We then created an average ΔF image consisting of 10 s frames corresponding to the maximum fluorescence interval (shown in Fig. 2c). This operation significantly reduced the fluorescence signals from spontaneous activity, which is typically short (lasting 1–2 s) and consists of small bursts, and enriched evoked activity, which is a more sustained (more than 10 s), larger rise in fluorescent intensity. The fluorescence traces of individual positive cells were further examined to confirm the sustained nature of the response. The number of activated cells was quantified using ImageJ. To quantify the number of viable cells during the imaging experiments, we counted the number of G-CaMP3-positive cells responsive to 40 mM KCl in mACSF.

Electrophysiology. Loose patch recordings were performed at room temperature with a Multiclamp 700B (Axon Instruments). Data were recorded at 10 kHz, low pass filtered at 2 kHz and digitized with a Digidata 1440A digitizer (Axon Instruments). Borosilicate glass (Sutter Instruments Co., o.d. 1.5 mm, i.d. 0.86 mm) patch pipettes (3–8 M Ω) were pulled on a Flaming/Brown micropipette puller (Sutter Instrument Co.). The same mACSF was used as for the pipette solution. Data were acquired with pClamp and analysed in Matlab. Pneumatic electronic valves (Clippard Instruments) were used to control the flow of stimuli. Electronic valves were controlled via digital output from the Digidata 1440 A digitizer. The valves were opened for 20 s in every stimulated trial. For rat stimulus, we used 1:200 dilution of rat urine (male CD rats, Charles River, 2–6 months old) in mACSF. For ferret stimuli, ~50 ml volume of ferret bedding containing urine, faeces, fur and gland excretions was extracted with 50 ml of mACSF overnight at 4 °C, then filtered and used for experiments.

31. Miller, R. A. *et al.* Mouse (*Mus musculus*) stocks derived from tropical islands: new models for genetic analysis of life-history traits. *J. Zool.* **250**, 95–104 (2000).
32. Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).
33. Danciger, E., Mettling, C., Vidal, M., Morris, R. & Margolis, F. Olfactory marker protein gene: its structure and olfactory neuron-specific expression in transgenic mice. *Proc. Natl Acad. Sci. USA* **86**, 8565–8569 (1989).

Theta-paced flickering between place-cell maps in the hippocampus

Karel Jezek^{1,2}, Espen J. Henriksen¹, Alessandro Treves^{1,3}, Edvard I. Moser¹ & May-Britt Moser¹

The ability to recall discrete memories is thought to depend on the formation of attractor states in recurrent neural networks^{1–4}. In such networks, representations can be reactivated reliably from subsets of the cues that were present when the memory was encoded, at the same time as interference from competing representations is minimized. Theoretical studies have pointed to the recurrent CA3 system of the hippocampus as a possible attractor network^{3,4}. Consistent with predictions from these studies, experiments have shown that place representations in CA3 and downstream CA1 tolerate small changes in the configuration of the environment but switch to uncorrelated representations when dissimilarities become larger^{5–9}. However, the kinetics supporting such network transitions, at the subsecond timescale, is poorly understood. Here we show in rats that instantaneous transformation of the spatial context does not change the hippocampal representation all at once but is followed by temporary bistability in the discharge activity of CA3 ensembles. Rather than sliding through a continuum of intermediate activity states, the CA3 network undergoes a short period of competitive flickering between preformed representations of the past and present environment before settling on the latter. Network flickers are extremely fast, often with complete replacement of the active ensemble from one theta cycle to the next. Within individual cycles, segregation is stronger towards the end, when firing starts to decline, pointing to the theta cycle as a temporal unit for expression of attractor states in the hippocampus. Repetition of pattern-completion processes across successive theta cycles may facilitate error correction and enhance discriminative power in the presence of weak and ambiguous input cues.

The place-cell population of the hippocampus is thought to create a neural representation of the spatial environment¹⁰. Accumulating evidence indicates that environments are generally represented in hippocampal cells by many discrete maps, each corresponding to a distinct environment or a unique experience within the environment^{5,9,11–13}. Which map is active at any given time depends on external sensory inputs as well as recent history^{7,14}. Incongruity between the active map and sensory inputs may lead to partial or complete replacement of the active representation^{15,16}. The kinetics of map substitutions has remained elusive owing to a shortage of experimental and analytical tools for subsecond-timescale neural population analyses.

Here, we developed tools to determine how local network activity evolves in the hippocampus in response to sudden changes in the cues that define spatial context. Rats with tetrodes in CA3 (Supplementary Fig. 1) were first trained on separate trials in two boxes with different sets of light cues in a dark room (boxes A and B). The procedure favours the development of uncorrelated place representations in A and B^{5,9} (Supplementary Fig. 2). After several days of training in each box, the rats started in one of the environments; then, after 40–60 s, the cues were switched instantaneously to those of the other environment, effectively ‘teleporting’ the rat from A to B or vice versa. A total of 169 such trials, hereafter referred to as teleportation trials, were performed,

with an average of 33 ± 3 active cells per day (mean \pm s.e.m.; total of 358 active cells; 11 days; 6 rats).

To examine the evolution of network activity after the cue change, we first established expected firing patterns for all locations in each environment. Firing rates were determined for each cell in each spatial bin of each box on separate reference trials (30×30 bins; Fig. 1a). As expected for CA3 place cells¹⁷, the subsets of active cells in the two boxes overlapped minimally: cells with activity in both A and B generally fired at unrelated positions (spatial correlation: 0.112 ± 0.019). The nearly orthogonal nature of the baseline representations made it possible, in principle, to infer from any subsequent cell sample which of the two environments was represented in the hippocampal network at the time of recording. Thus, in the next step, we compared the evolution of activity over successive theta cycles in the teleportation test with activity at corresponding locations in the reference environments. The theta phase with the lowest overall firing rate was used to segment the recording from each teleportation trial into individual theta cycles (Fig. 1b, c and Supplementary Fig. 3). A population vector was then constructed for each theta cycle, consisting of the number of spikes of each of the C simultaneously recorded cells (Fig. 1c). For each theta cycle, the population vector was correlated with vectors of average firing rates (in Hz) for the same location in each reference recording. We first used a simple linear measure of correlation, the un-normalized dot product (DP), between the test population vector \mathbf{r} and each of the two reference vectors, \mathbf{r}_A and \mathbf{r}_B . If \mathbf{r}_A and \mathbf{r}_B were strictly orthogonal and after teleportation test vectors were to be linear combinations $\mathbf{r} \approx c_A \mathbf{r}_A + c_B \mathbf{r}_B$, the DP measure would allow a direct estimate of the coefficients $c_A(t)$ and $c_B(t)$, that is, of the strength with which the two representations are relayed to downstream neurons at time t ^{2,4}. Only trials with continuous theta activity were analysed (149 out of 169 teleportation trials).

The switching of spatial cues was followed by a characteristic pattern of network activity (Fig. 2). Before the transformation, most theta cycles correlated strongly with reference vectors for the first environment (denoted as I, whether A or B) but not for the second (II), as expected (Figs 1d and 2a, b). After cue switching, most theta cycles correlated with the reference vectors for II but the magnitude of the dot product was variable and the network occasionally ‘flickered’ back to strong correlations with I. To investigate the statistics of this flickering, we plotted for each theta cycle the correlation of the momentary population vector with the reference vectors from the same environment (x -axis) and the alternative environment (y -axis) (Fig. 2c). The analysis confirmed that cue switching increased the proportion of cycles correlated with the alternative environment but also identified a subset of cycles that correlated modestly with both environments at the same time. This led us to ask if the instantaneous activity during the transition reflected a simple linear combination of fragments of the A and B representations, as would be expected in the absence of population coherence. We consequently determined whether, in the pooled data, the proportion of theta cycles correlating with both environments at

¹Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, Olav Kyrres gate 9, MTF5, 7489 Trondheim, Norway. ²Department of Physiology of Memory, Institute of Physiology, Videnska 1083, Academy of Sciences of the Czech Republic, 142 20 Prague 4-Krc, Czech Republic. ³Cognitive Neuroscience Sector, SISSA International School for Advanced Studies, via Bonomea 265, 34136 Trieste, Italy.

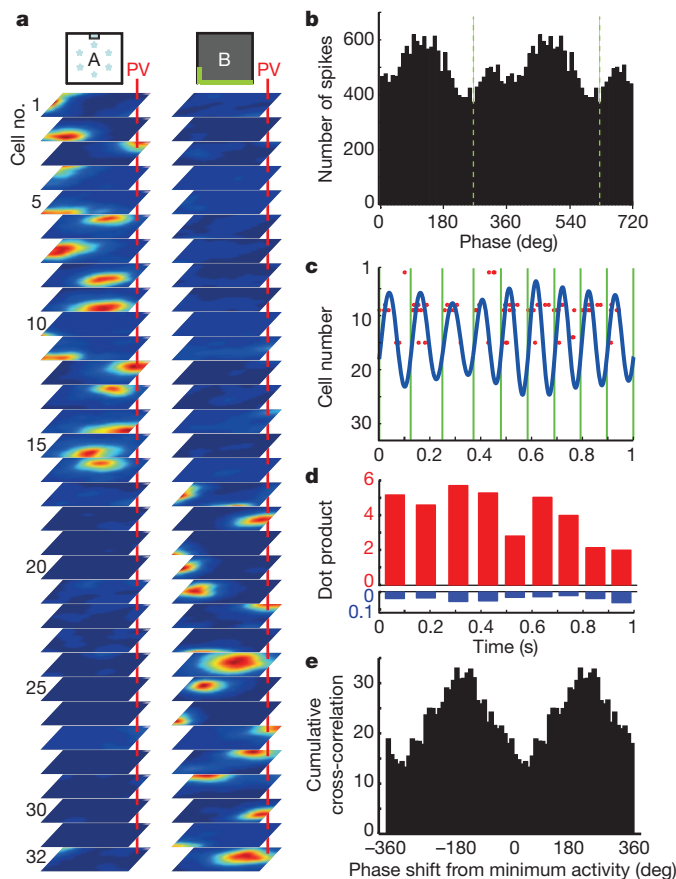


Figure 1 | Procedures for analysing hippocampal transition dynamics.

a, Stack of firing-rate maps in box A (white floor lights; left) and box B (green wall lights; right) for an example set of 32 simultaneously recorded hippocampal CA3 units. Each map shows a colour-coded distribution of firing rates across the square test box (blue, silent; red, maximum). Red line, one of 30×30 population vectors (PV) constructed from the activity of the entire cell ensemble in a given 2×2 cm position bin. Note strong difference in population vectors for A and B. **b**, Theta phase modulation for all pyramidal cells on a representative trial. Spike number is shown as a function of theta phase (bin size 10 deg). Dashed green line, phase with lowest firing rate, used to define boundary between cycles. **c**, Representative spike distribution across theta cycles in the stable state. Rasters of red dots show spike times of individual cells in relation to 6–11 Hz filtered local electroencephalogram (EEG; blue). Green lines indicate theta-cycle boundaries (**b**). The ensemble distribution of activity during one cycle represents the momentary population vector. **d**, Dot-product correlation between momentary population vector and reference vectors at the corresponding position in A (red) and B (blue) during a baseline trial in A. **e**, Cumulative product between correlations with each of the reference environments as a function of the phase for segmentation of theta cycles (0, phase of minimum activity).

the same time ($\mathbf{r} \cdot \mathbf{r}_A > C$, $\mathbf{r} \cdot \mathbf{r}_B > C$, referred to as ‘mixed theta cycles’) was lower than expected if each single unit expressed either one or the other representation independently of the other units. A total of 1.25% of the theta cycles in the recorded data were mixed. This number was lower than in 970 out of 1,000 randomly recombined (shuffled) population vectors (that is, $P < 0.03$; Fig. 2d). The separation between the A- and B-correlated representations was strongest when the cycles were chunked at the point of the lowest average firing rate in the population (Fig. 1b, c, e and Supplementary Fig. 3). Thus, mixed representations existed but were rare. Transitions between orthogonal maps tend to occur in an all-or-none manner, with the entire network flickering coherently at time scales of approximately a tenth of a second.

We subsequently examined the evolution of network activity within the theta cycle. Each cycle was divided into two halves, and mixed states were defined as those half-cycles for which $\mathbf{r} \cdot \mathbf{r}_A > C/2$, $\mathbf{r} \cdot \mathbf{r}_B > C/2$. Mixed population vectors were less abundant than

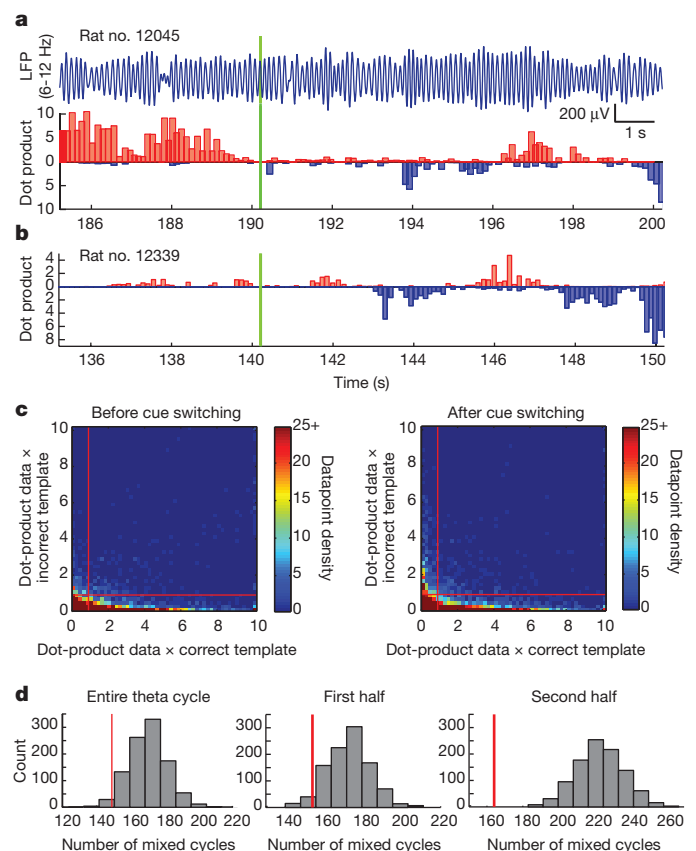


Figure 2 | Theta cycles correlate with either of the reference environments but rarely with both simultaneously.

a, Top, local hippocampal EEG during spatial cue switching from A to B (filtered at 6–11 Hz). Bottom, dot-product correlation between momentary population vectors and reference vectors from A (red) and B (blue) for successive theta cycles before and after cue switching. Dot products are un-normalized (just divided by the number C of recorded cells). All correlations are positive but, for clarity, A and B correlations are plotted in opposite directions. Green line indicates light switch. EEG and ensemble activity were sampled simultaneously. Note that ensemble activity flickered back to the A representation several times after cue switching. Note also the variation in the dot product. **b**, Another example of network flickering induced by switching of spatial cues. **c**, Matrices reporting the number of cycles falling in each 0.2×0.2 bin of the dot-product correlations $\mathbf{r} \cdot \mathbf{r}_{A,B}/C$ between momentary population vectors and reference vectors for the present environment (x -axis) or the alternative environment (y -axis). Left, before cue switching. Right, after cue switching (starting from the first cycle correlated with new environment). Note that mixed cycles, defined as cycles with both $\mathbf{r} \cdot \mathbf{r}_A$ and $\mathbf{r} \cdot \mathbf{r}_B$ exceeding C ($x > 1$, $y > 1$; indicated by red lines), were rare. **d**, Histograms showing that the number of mixed states after cue switching (red line) is lower than expected from shuffled versions of the same data ($n = 1,000$; grey histogram). Note that mixed states became less frequent during the second half of the theta cycle.

expected from shuffled data during both half-cycles; however, in the shuffled data, the frequency of recombinations with more mixed population vectors than in the observed data increased from 958/1,000 during the first half ($P < 0.05$) to 1,000/1,000 during the second ($P < 0.001$) (Fig. 2d). The low incidence of mixed population vectors at the end of the theta cycle suggests that representations evolve from partially segregated to fully segregated within each activity period.

We then asked how A-correlated and B-correlated theta cycles were organized in time. Because dot products can vary from 0 to indefinitely large, we switched to Pearson product-moment correlations, which by normalizing the correlations to within a fixed $[-1, +1]$ range allow successive theta cycles to be compared more directly (Fig. 3a–c and Supplementary Figs 4–6). Individual theta cycles were now only considered if at least two cells were active (for higher thresholds, see Supplementary Fig. 7). As observed with the dot products, the

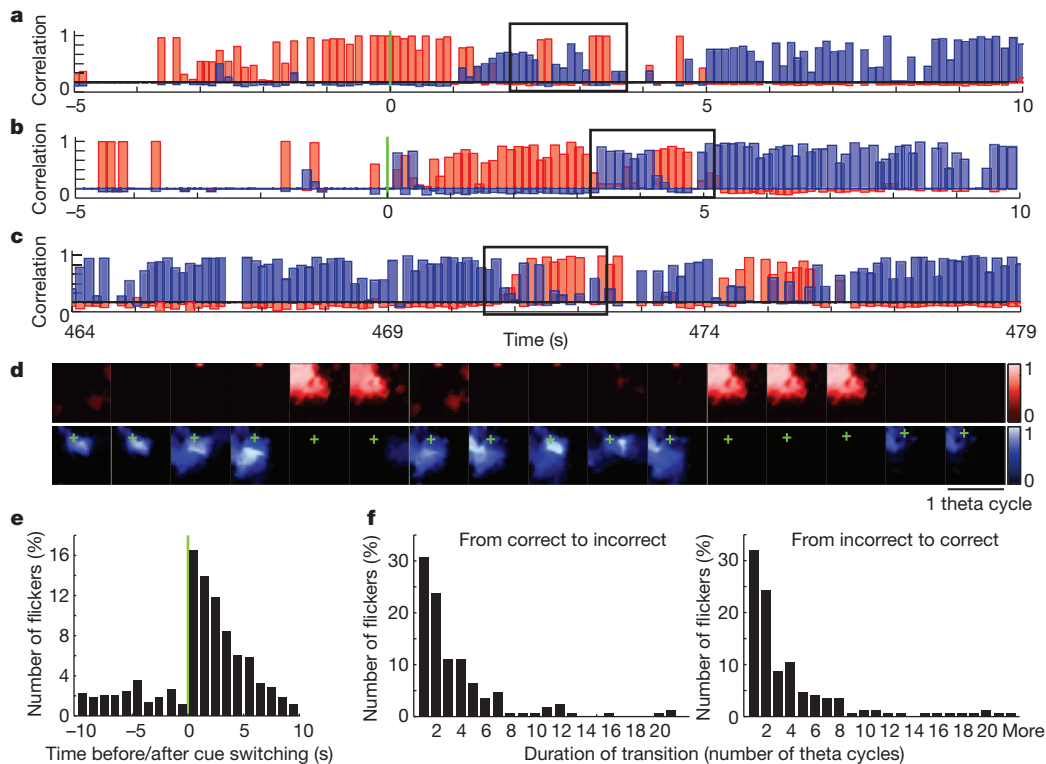


Figure 3 | Temporal dynamics of network flickering. **a–c**, Pearson product-moment correlations showing evolution of population vector correlations after spatial cue switching from box I to II (red, correlation with I; blue, with II; green line, cue switch). Note frequent flickers to the original representation after cue switching in **a** and **b**. **c**, Spontaneous flickering between cue-switch events. **d**, Spatial distribution of correlations between momentary population vectors and reference vectors in the framed area in **a**. Each row shows correlation matrices

momentary population vectors correlated strongly with either A or B but rarely with both. Before cue switching, theta cycles were nearly exclusively correlated with reference vectors for the initial environment (I). After cue switching, the network switched almost instantaneously to high correlation with reference vectors for II but then relapsed to I several times during the subsequent seconds (Fig. 3a, b and Supplementary Figs 5 and 6) and occasionally tens of seconds after cue switching (Fig. 3c). These relapses, or flickers, were confined to discrete periods of one or several theta cycles. When the population vector was correlated with reference vectors from other locations in the represented environment, the correlation generally decreased with distance from the animal (Fig. 3d and Supplementary Fig. 8).

To quantify the frequency, timing and duration of flicker events, we defined individual theta cycles as A-correlated if the correlation with reference environment A was above the 95th percentile for $B \times A$ correlations in the reference sessions (that is, more similar to A than 95% of theta cycles in B) and if the correlation with reference environment B was simultaneously below the 5th percentile for $B \times B$ (that is, different from most theta cycles in B) (Supplementary Figs 9 and 10). The analysis showed a clear increase in the frequency of network flickers during the first seconds following the cue switch from I to II, after the network had switched to the II representation for the first time (generally 0–1 s after the cue change; Supplementary Fig. 11 and Supplementary Table 1). The fraction of theta cycles participating in flicker episodes, estimated with Pearson correlations, increased from a stable baseline of 1–3% before cue switching to a level of 10–15% during the first 5 s after the first network switch (Fig. 3e and Supplementary Fig. 12), confirming the tendency revealed by the dot-product analyses (Fig. 2c). The interquartile range of flicker durations increased from 1–1 theta cycles before cue switching to 1–4 during the first 10 s after cue switching (Wilcoxon rank-sum test:

for 16 consecutive theta cycles. Top, correlation with I; bottom, with II. Correlation is colour coded (scale bar). + indicates rat position. **e**, Percentage of flickers to the alternative representation as a function of time before and after cue switching. **f**, Distribution plots showing the time for the network to switch from present ('correct') to past ('incorrect') representation, or vice versa, in number of theta cycles. One cycle means that cycles with alternative representations were consecutive. Note the predominance of immediate transitions.

$Z = 2.27$, $P < 0.03$; Supplementary Fig. 11b and Supplementary Table 1). Flicker events were distributed across the entire recording box (Supplementary Fig. 13) and showed no preference for running or heading direction (Supplementary Fig. 14). Flickering was apparent also in CA1 but discrete relapses were clearly less frequent, possibly due to the less reverberating architecture of this subfield (Supplementary Fig. 15).

Transitions between representations occurred within less than a single theta cycle. After the last cycle in a series of non-flicker theta cycles, the alternative representation was fully expressed already in the subsequent theta period in 30.8% of the flicker events where flicker duration could be determined (Fig. 3f). The corresponding percentage of single-cycle returns to the 'correct' representation was 32.0%. In 57.0% of transitions from 'correct' to 'incorrect' reference frame and 62.2% of transitions back from 'incorrect' to 'correct', the time course remained undetermined because the intervening theta cycle contained fewer than two spikes. In each instance of an immediate transition, the network representation was fully developed from the outset, that is, the correlation with the new environment did not increase further within the flicker period ($r = -0.06 \pm 0.11$).

To determine further if the flicker episodes were patterned by the theta oscillation, we finally compared the transition dynamics of our theta-based segmentation procedure with segmentations based on fixed time bins of different width (range 44–500 ms; Supplementary Fig. 3). Among the fixed bins, direct transitions were most abundant when the bins matched the average duration of theta cycles (125 versus 120.4 ms, respectively). The abundance of sharp transitions increased further when the trial was segmented by actual rather than fixed theta periods, with cycles split at the phase with the minimum firing rate (Fig. 1b, c, e and Supplementary Fig. 3). Collectively, these observations suggest strongly that the transitions were paced by the theta rhythm.

Our study provides evidence for competitive interactions between hippocampal representations during changes in spatial reference frame. Although a small subset of the population vectors correlated with both reference environments after the cue change, the number of such mixed states was lower than expected from a sample of independent single units, especially during the second half of each theta cycle. In most cases, the network either switched all at once or flickered between mutually exclusive representations until, after a few seconds, it settled in one of the alternatives. The sustained separation of the neural activity pattern is consistent with the notion that spatial environments, as a whole, are stored as discrete attractors in neural networks of the hippocampus or associated areas such as the entorhinal cortex^{1–4,8,9}. These discontinuities differ from the continuous or quasi-continuous nature of spatial maps for individual environments^{12,18–20}, where sweeps can follow unbroken trajectories, even when retrieval occurs in the absence of actual movement²¹. Changes in attention or experience are likely to generate continuous transitions of the latter type more or less constantly in all environments^{7,22–26}. The low frequency of flickering in the baseline state of the teleportation task suggests, however, that switches to uncorrelated attractor maps are rare and occur primarily when cues are ambiguous or in conflict across sensory modalities.

The time course of flickering episodes has implications for the mechanisms of ensemble activation. It took often only a single theta cycle to fully reactivate a pre-established representation during a flicker event, consistent with models of theta phase precession in which ensembles are activated by propagation through recurrent collaterals after afferent input to a subset of cells early in the theta cycle²⁷. The idea that sensory influences can override attractors at the beginning of the theta cycle, whereas subsequent activity is determined more exclusively by propagation through associative connections, receives further support from the fact that the small number of mixed population vectors occurred primarily during the first half of the theta cycle. The recreation of spatial representations on successive theta cycles in the hippocampus is fundamentally different from the pattern-completion dynamics observed, for example, in the inferior temporal cortex, where the activation may proceed only once per stimulus presentation, at a timescale short enough to facilitate perception but with little opportunity for error correction²⁸. In the hippocampus, repeated convergence to an attractor state might allow the system to self-correct and thereby enhance its discriminative power under conditions where input cues are weak and ambiguous. The present data point to theta cycles as organizational units for this repetitive process but do not preclude additional structure at faster time scales, for example with gamma cycles as units for cell assembly sequences within the theta cycle^{29,30}.

METHODS SUMMARY

Neuronal ensemble activity was recorded from ensembles of CA3 cells while rats foraged in either of two distinct enclosures in a dark room. The enclosures were identical except for internal lights. Testing began by placing the rat in one of the two boxes. After 40–60 s, the light cues were switched, effectively ‘teleporting’ the animal to the other environment. To examine the evolution of network activity, population vectors were defined from the firing rates of all simultaneously recorded cells for every theta cycle before and after the switching of spatial cues. Each population vector was correlated with reference vectors defined from the activity of the same cells at the same spatial location in each box on separate baseline trials.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 November 2010; accepted 10 August 2011.

Published online 28 September 2011.

1. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
2. Amit, D. J., Gutfreund, H. & Sompolinsky, H. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.* **55**, 1530–1533 (1985).
3. McNaughton, B. L. & Morris, R. G. M. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415 (1987).

4. Treves, A. & Rolls, E. T. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2**, 189–199 (1992).
5. Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.* **7**, 1951–1968 (1987).
6. Lee, I., Yoganarasimha, D., Rao, G. & Knierim, J. J. Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature* **430**, 456–459 (2004).
7. Leutgeb, J. K. *et al.* Progressive transformation of hippocampal neuronal representations in ‘morphed’ environments. *Neuron* **48**, 345–358 (2005).
8. Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O’Keefe, J. Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005).
9. Colgin, L. L. *et al.* Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *J. Neurophysiol.* **104**, 35–50 (2010).
10. O’Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
11. Gothard, K. M., Skaggs, W. E., Moore, K. M. & McNaughton, B. L. Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *J. Neurosci.* **16**, 823–835 (1996).
12. Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).
13. Derdikman, D. *et al.* Fragmentation of grid cell maps in a multicompartment environment. *Nature Neurosci.* **12**, 1325–1332 (2009).
14. O’Keefe, J. & Speakman, A. Single unit activity in the rat during a spatial memory task. *Exp. Brain Res.* **68**, 1–27 (1987).
15. Gothard, K. M., Skaggs, W. E. & McNaughton, B. L. Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *J. Neurosci.* **16**, 8027–8040 (1996).
16. Skaggs, W. E. & McNaughton, B. L. Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *J. Neurosci.* **18**, 8455–8466 (1998).
17. Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B. & Moser, E. I. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* **305**, 1295–1298 (2004).
18. Tsodyks, M. & Sejnowski, T. Associative memory and hippocampal place cells. *Int. J. Neural Syst.* **6**, 81–86 (1995).
19. Romani, S. & Tsodyks, M. Continuous attractors with morphed/correlated maps. *PLoS Comput. Biol.* **6**, e1000869 (2010).
20. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M. B. Path integration and the neural basis of the ‘cognitive map’. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
21. Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189 (2007).
22. Fenton, A. A. & Muller, R. U. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc. Natl Acad. Sci. USA* **95**, 3182–3187 (1998).
23. Olypher, A. V., Lánský, P. & Fenton, A. A. Properties of the extra-positional signal in hippocampal place cell discharge derived from the overdispersion in location-specific firing. *Neuroscience* **111**, 553–566 (2002).
24. Jackson, J. & Redish, A. D. Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. *Hippocampus* **17**, 1209–1229 (2007).
25. Kelemen, E. & Fenton, A. A. Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. *PLoS Biol.* **8**, e1000403 (2010).
26. Blumenfeld, B., Preminger, S., Sagi, D. & Tsodyks, M. Dynamics of memory representations in networks with novelty-facilitated synaptic plasticity. *Neuron* **52**, 383–394 (2006).
27. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* **6**, 271–280 (1996).
28. Akrami, A., Liu, Y., Treves, A. & Jagadeesh, B. Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. *Cereb. Cortex* **19**, 760–776 (2009).
29. Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of cell assemblies in the hippocampus. *Nature* **424**, 552–556 (2003).
30. Colgin, L. L. *et al.* Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* **462**, 353–357 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Skjerpeng, A. M. Amundsgård, K. Haugen, K. Jenssen, E. Kråkvik and H. Waade for technical assistance. The work was supported by the 7th Framework Programme of the European Commission (‘SPACEBRAIN’, grant agreement no. 200873), an Advanced Investigator Grant to M.-B.M. from the European Research Council (‘ENSEMBLE’), the Kavli Foundation, a Centre of Excellence grant from the Norwegian Research Council, and research projects MSM CR LC554, 1M0517 and AV0250110509 at the Academy of Sciences of the Czech Republic.

Author Contributions K.J., A.T., M.-B.M. and E.I.M. designed the study and discussed analyses and results; K.J. built the apparatus, K.J. and E.J.H. performed experiments, K.J. performed analyses; E.I.M. wrote the paper with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.J. (karel.jezek@biomed.cas.cz) or E.I.M. (edvard.moser@ntnu.no).

METHODS

Subjects. Six male Long Evans rats (400–500 g at implantation) were housed individually in transparent Plexiglass cages (45 cm × 30 cm × 35 cm). The animals were kept at ~90% of their initial free-feeding body weight and maintained on a 12-h light/12-h dark schedule. All testing occurred in the dark phase. The experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes.

Electrode preparation and surgery. Neuronal ensemble activity was recorded from ensembles of CA3 or CA1 cells in rats implanted with a 'hyperdrive' containing 14 independently movable tetrodes assembled in a circular bundle. Tetrodes were twisted from four 17- μ m polyimide-coated platinum-iridium wires (90% and 10%, respectively; California Fine Wire Company). Electrode tips were plated with platinum to reduce electrode impedances to 120–200 k Ω at 1 kHz.

The animals were food deprived 12 h before surgery started. In four animals, anaesthesia was induced by first placing the animal in a closed glass box filled with isoflurane vapour and then giving the animal an intraperitoneal injection of Equithesin (pentobarbital and chloral hydrate; 1.0 ml per 250 g body weight). Two animals (15272 and 15273) were anaesthetized with isoflurane (induction chamber level of 4.0% while the rats were secured in the stereotaxic apparatus, with an air flow at 1,400 ml min⁻¹; isoflurane was then gradually reduced to 1–2% during the course of the surgery). Supplementary anaesthesia was given when breathing and reflexes changed. Local anaesthetic (Xylocain) was applied on the skin before making the incision. The hyperdrive was then implanted. The tetrodes were inserted above CA3 of the right hippocampus, with the centre of the bundle at anterior–posterior 3.8 mm and medial–lateral 3.0 mm relative to bregma. Jewellers' screws and dental cement were used to secure the hyperdrive to the skull. Two screws were connected to hyperdrive ground. All tetrodes were turned after the surgery to be sure they were in the brain.

Tetrode positions. Over the course of ~3–4 weeks, the majority of the tetrodes were lowered towards CA3 in steps of 50 μ m or less while the rat rested on a towel in a large flower pot on a pedestal. Turning was slowed down when large-amplitude theta-modulated complex-spike activity appeared in CA3 at depths of approximately 3.0 mm. The tetrode depths were tweaked to get the maximal number of simultaneously recorded CA3 cells at the start of cue switching. To maintain stable recordings, the electrodes were not moved at all on the day of recording. A few tetrodes were left in CA1; data from these tetrodes were analysed separately. Two of the tetrodes were used, respectively, to record a reference signal from the corpus callosum and an EEG signal from the stratum lacunosum-moleculare.

Behavioural training procedures. The rats were trained to collect food morsels in either of two distinct 60 cm × 60 cm enclosures with 40-cm walls located in a dark curtained environment (Supplementary Fig. 2). The boxes rested on a plexiglass plate fixed 10 cm above the floor of the room. The boxes were identical except for the arrangement of a number of internal lights. Beneath the plexiglass plate of box A there was a panel of eight light-emitting white diodes (LEDs) organized into a circle (50 cm in diameter) and placed centrally under the plexiglass plate to be visible through the floor. The box was polarized by another LED at the upper edge of one of the walls. Box B was illuminated by a 60-cm-long array of green LEDs lining 40 cm of the upper edge of the wall opposite to the directional LED in A and 20 cm of one of the adjacent walls (Supplementary Fig. 13a). The LEDs were the only light source in the room.

Training occurred in four stages. During stage 1, the boxes were located next to each other, connected by a 20 × 20 cm (width × length) passageway that allowed the rat to shuttle between the boxes in order to associate each box with a different set of path integrator coordinates⁹. The rat was permitted to travel freely between the boxes for 20 min on at least three trials. Trials were separated by 20 min intervals, during which the animal rested on a towel in a pedestal outside of the curtains. At stage 2, the corridor was removed and the animal explored the boxes individually on alternating trials (3 trials in each). At stage 3, the boxes were replaced by a single box made of the same material and equipped with both sets of lights. The box was placed on alternating trials at the two original locations. When presented at the original position of box A, the set of LEDs defining A was switched on; when in the place of box B, the respective lights of B were active instead. Again, the rat was tested 3 times in each environment on alternating occasions. Finally, at Stage 4, the box was moved to a central location between the two original box locations. The animal received alternating trials with each set of lights (2 consecutive days, each day 3 pairs of trials). During all stages, at the start of each trial, the rat was taken from the flower pot on the pedestal outside of the curtains and placed, without disorientation, into the environment with the eyes gently covered by the experimenter's palm. Between trials, the rat rested for 20 min in the flower pot. During this period, the boxes were thoroughly washed with a wet tissue and dried.

On the test day, the rat started with a 20 min rest trial in the flower pot. The animal was then tested for 10 min in each box configuration (A and B, respectively). Then, during the third trial, after 40–60 s of baseline recording in one of the configurations (for example, A), the lights were switched instantaneously to the other configuration (for example, B). Additional 'teleportations' were performed every subsequent 40–60 s until 10 min had passed. A resting trial in the flower pot was recorded at the end. The electrodes were then lowered deeper into the CA3 and the experiment was repeated on the subsequent day in those cases where spikes with sufficient amplitudes from new cells could be recorded.

Running was motivated by small crumbles of cookies thrown into the box at 10–20 s intervals. Three types of cookies were used: vanilla, chocolate and unflavoured. Vanilla and unflavoured were given in configuration A, chocolate and unflavoured in configuration B. The proportion between flavoured and unflavoured cookies was approximately 1:1 during stage 1 and 1:4 during the following stages. During teleportation trials, only unflavoured crumbles were offered.

Population vector analyses. Theta waves were identified from filtered local EEG traces as described in the Supplementary Methods. The evolution of unit activity over successive theta cycles before and after cue switching was estimated by defining a population vector for each theta cycle, consisting of the number of spikes fired by each cell in the array, and comparing it to a population vector for the same cells at the corresponding location in each of the reference environments (Fig. 1a, c). Boundaries between theta cycles were determined by plotting, for all cells on the entire teleportation session, the number of spikes as a function of theta phase (Fig. 1b). The boundary between successive theta cycles was then defined by the theta phase with the lowest overall firing rate (bins of 10 deg; Fig. 1b, c). For each theta cycle, the momentary population vector was correlated with reference population vectors for the same cell sample at the same location in each of the reference environments. The reference vector (expressed in Hz rather than as a spike count, normalized by the time spent at each location) was based on activity across the entire reference trial for each of the environments on the test day. In a subset of the analyses, the momentary population vector was compared with all reference population vectors in the two boxes, that is, not only those corresponding to the animal's current location. Correlations between momentary population vectors and reference vectors were quantified by dot products as well as Pearson product–moment correlations. Dot products were calculated for all theta cycles, including those with no activity, where the dot product is zero. Pearson correlations, which cannot be defined for theta cycles with no spikes, were calculated by excluding also theta cycles with a single unit active, that is, these correlations were based on the subset of theta cycles that included at least two active cells. In a subset of the analyses, also cycles with only two active units were excluded, and in a further control, also those with only three active units were discarded (Supplementary Fig. 7).

To determine if the similarity between momentary population vectors during cue switching and population vectors in the reference environment was larger than expected by chance, we correlated population vectors from successive theta cycles in reference recordings from A and B with mean population vectors generated from activity in the same position across the entire session in the same environment or the alternative environment. When referenced to the same environment ($A \times A$ or $B \times B$), the distribution of Pearson correlations was centred at high correlation values with a long tail of low values (Supplementary Fig. 9, top row). When referenced to the alternative environment ($A \times B$ or $B \times A$), the distributions were centred at slightly negative values with a long tail towards infrequent high positive values (Supplementary Fig. 9, bottom row). We then determined 5th and 95th percentile values for each distribution and used these as criteria to identify flickers to the alternative representation. Individual theta cycles were defined as A-correlated if the correlation with reference environment A was above the 95th percentile for $B \times A$ (that is, more similar to A than 95% of the theta cycles in B) and if the correlation with reference environment B was below the 5th percentile for $B \times B$ (that is, more different from the reference in B than most theta cycles in B). Conversely, theta cycles were defined as B-correlated if the correlation with reference environment B exceeded the 95th percentile for $A \times B$ and if the correlation with reference environment A was lower than the 5th percentile for $A \times A$. In separate analyses, the momentary population vectors were compared not only with reference vectors at the same location but at all 900 bin locations in the box (Supplementary Fig. 8).

The abundance of theta cycles that correlate with both reference environments ('mixed' theta cycles) was determined by comparing the data with spike patterns obtained by 'shuffling' the activity of individual units. For each theta cycle during the 10 s preceding and succeeding the cue switching, the number of spikes produced by each unit recorded during the cue switching was drawn at random from among all theta cycles recorded in the same physical location in the same environment during the corresponding period (most of these theta cycles were correlated with the current environment, some with the alternative one). The analysis was

limited to theta cycles for which a minimum of 3 theta cycles had been recorded in that particular location in the relevant period, with 'same' location defined as the same $6\text{ cm} \times 6\text{ cm}$ spatial bin of the same recording box. The shuffled population vectors can be conceived as approximating the linear combination $\mathbf{r}_S \approx \mathbf{A} \cdot \mathbf{r}_A + \mathbf{B} \cdot \mathbf{r}_B$ where \mathbf{A} and $\mathbf{B} = 1 - \mathbf{A}$ are random binary vectors (for example, (0,1,0,0,1,0,1,1,...)) indicating whether each unit was drawn from the A or B representation of that location. If the current environment is A, a particular theta cycle in the original data can express activity close to the reference vector A in that location, or to the reference vector B (for example, during a flicker event) or to a mixture of the two, or just noise or, rather frequently, no activity at all. The abundance of mixed theta cycles was determined by the un-normalized dot product DP of the population vector in each theta cycle with the reference vectors in A and B, $DP_A = \mathbf{r} \cdot \mathbf{r}_A / C$ and $DP_B = \mathbf{r} \cdot \mathbf{r}_B / C$ (just divided by the number C of simultaneously recorded units). DP values range from 0 to about 10 (note that they are expressed in Hz, as the momentary population vector includes spike counts whereas the reference vector indicates firing rates), and very rarely beyond, but are mostly clustered close to zero because of the presence of theta cycles with no

spikes or with, for example, a single spike from a unit that emits on average less than 1 spike in that physical location in both reference vectors. We arbitrarily set $DP > 1$ as the criterion for substantial correlation, and $DP > 0.5$ for half-cycles.

The speed of instantaneous transitions from one representation (for example, A) to the other (for example, B) was estimated by cross-correlating the sequence of correlations with reference environment A and the sequence of correlations with B one time-bin later, excluding, in this analysis, units with overlapping fields in the two environments and trials with less than 20 non-overlapping units (2 out of 11; Supplementary Fig. 3). Bin widths were then varied from 44 ms to 500 ms to determine the transition time that gave the largest cross-correlation. Bins were also defined by actual theta cycles, with separations at the point of minimum spike activity in the population (Fig. 1b, c) and at other phases 10 deg apart (Supplementary Fig. 3). Cross-correlation values were corrected for overlap between A and B representations by subtracting the cross-correlation at zero time lag.

Other. Recording procedures, criteria for spike sorting, construction of rate maps, analysis of theta rhythm and histological procedures are described in Supplementary Materials and Methods.

Peripheral education of the immune system by colonic commensal microbiota

Stephanie K. Lathrop^{1†}, Seth M. Bloom², Sindhuja M. Rao¹, Katherine Nutsch¹, Chan-Wang Lio¹, Nicole Santacruz¹, Daniel A. Peterson^{3†}, Thaddeus S. Stappenbeck² & Chyi-Song Hsieh¹

The instruction of the immune system to be tolerant of self, thereby preventing autoimmunity, is facilitated by the education of T cells in a specialized organ, the thymus, in which self-reactive cells are either eliminated or differentiated into tolerogenic Foxp3⁺ regulatory T (T_{reg}) cells¹. However, it is unknown whether T cells are also educated to be tolerant of foreign antigens, such as those from commensal bacteria, to prevent immunopathology such as inflammatory bowel disease^{2–4}. Here we show that encounter with commensal microbiota results in the peripheral generation of T_{reg} cells rather than pathogenic effectors. We observed that colonic T_{reg} cells used T-cell antigen receptors (TCRs) different from those used by T_{reg} cells in other locations, implying an important role for local antigens in shaping the colonic T_{reg}-cell population. Many of the local antigens seemed to be derived from commensal bacteria, on the basis of the *in vitro* reactivity of common colon T_{reg} TCRs. These TCRs did not facilitate thymic T_{reg}-cell development, implying that many colonic T_{reg} cells arise instead by means of antigen-driven peripheral T_{reg}-cell development. Further analysis of two of these TCRs by the creation of retroviral bone marrow chimaeras and a TCR transgenic line revealed that microbiota indigenous to our mouse colony was required for the generation of colonic T_{reg} cells from otherwise naive T cells. If T cells expressing these TCRs fail to undergo T_{reg}-cell development and instead become effector cells, they have the potential to induce colitis, as evidenced by adoptive transfer studies. These results suggest that the efficient peripheral generation of antigen-specific populations of T_{reg} cells in response to an individual's microbiota provides important post-thymic education of the immune system to foreign antigens, thereby providing tolerance to commensal microbiota.

Although T_{reg} cells are required for the maintenance of gut tolerance³, commensal bacteria are not necessary for colonic T_{reg}-cell generation^{5,6} (Supplementary Fig. 1). Moreover, T_{reg} cells from germ-free mice can protect against colitis⁷. In contrast, extrathymic generation of T_{reg} cells that respond to foreign antigens has been demonstrated with TCR transgenic models of oral tolerance^{8,9}. Peripheral T_{reg}-cell development is also increased in the gut^{9,10}, a response that is potentially related to the presence of specialized antigen-presenting cells^{9,11–13}. Finally, species in the *Clostridium* genus¹⁴, and *Bacteroides fragilis*⁶ by means of a protease-resistant capsular polysaccharide, can increase the frequency or function of colonic T_{reg} cells, but may do so through innate immune receptors¹⁵. Thus, it remains unclear whether the protective colonic T_{reg}-cell population is generated against self antigens or foreign antigens derived from the commensal bacteria found in each individual.

Although TCR transgenic lines that respond to antigens derived from commensal bacteria have been described¹⁶, the normal *in vivo* frequency of those TCRs in the T_{reg}-cell subset as compared with the effector T-cell subsets is unknown. To study the TCRs normally found in

the colonic T_{reg}-cell population, we analysed the colonic TCR repertoire. As a result of the great diversity of the fully polyclonal TCR repertoire, we and others have used genetically engineered mice with limited polyclonal repertoires^{17–19}. The analysis of TCR α -chain repertoires of CD4⁺ T cells from the colonic lamina propria of mice expressing a fixed transgenic TCR β chain revealed that Foxp3⁺ T_{reg} cells use TCRs that are quite distinct from those of effector/memory (CD44^{hi}) and naive (CD44^{lo}) Foxp3[–] cells (Fig. 1a, b and Supplementary Figs 2 and 3). This is illustrated by using the Morisita–Horn similarity index (Fig. 1a), in which values from 0 to 1 represent low to high similarity between two data sets, and also by the analysis of the relative frequencies of individual common TCRs in each T-cell subset (Fig. 1b). Moreover, the analysis of TCR α chains of cells pooled from the secondary lymphoid organs and colons of additional mice showed that T_{reg} TCR usage in the colon differed greatly from that in the other organs (Fig. 1c, d and Supplementary Fig. 3). Like the T_{reg}-cell population, the effector/memory T-cell population expressed TCRs largely unique to the colon; however, these two subsets showed very little overlap. Thus, consistent with our previous observations in other peripheral locations²⁰, these TCR repertoire data suggest that the colonic T_{reg}-cell population is strongly shaped by the local antigenic milieu.

To assess whether the local antigens were bacterial in origin, we expressed colon T_{reg} TCRs (Supplementary Fig. 3) in a hybridoma cell line that contains a green fluorescent protein (GFP) reporter for nuclear factor of activated T cells (NFAT) activation as a readout for TCR engagement²¹. We initially screened these hybridomas against autoclaved colonic contents and were surprised to find that many (five of eight) showed some degree of reactivity to preparations from conventionally housed, but not germ-free, mice (Fig. 2a, Table 1 and Supplementary Fig. 4a, b). Colonic contents from Jackson Laboratories-sourced (Jax) *Rag1*^{–/–} mice were not recognized by four of these colonic T_{reg} TCRs unless they were first housed together with mice from our colony. In contrast to the colonic T_{reg} TCRs, none of the eight abundant colonic activated/memory (CD44^{hi}) TCRs (Supplementary Fig. 5), or of four other TCRs tested (including B8 and TR5 (ref. 20)) showed any reactivity (not shown). In addition, TCR recognition did not occur in the absence of dendritic cells and was blocked by antibodies against MHC class II (not shown). Thus, these data suggest that the antigens responsible for TCR activation are derived from microbes that can be passed between mice that are housed together.

We therefore attempted to identify bacteria recognized by these colonic T_{reg} TCRs by screening small pools of two or three heat-killed bacteria isolated in pure culture from the colonic contents of mice in our colony (Fig. 2b and Supplementary Figs 4c and 6). Two colonic T_{reg} TCRs (CT6 and CT7) reacted to one or more pools. Testing of individual isolates from these pools revealed that CT6 reacted with isolate ACNA18.1, identified by 16S ribosomal RNA gene sequencing as a previously uncharacterized *Clostridiales* species. All three isolates

¹Department of Medicine, Division of Rheumatology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA. [†]Present addresses: Laboratory of Intracellular Parasites, Rocky Mountain Laboratories, NIAID, NIH, Hamilton, Montana 59840, USA (S.K.L.); Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA (D.A.P.).

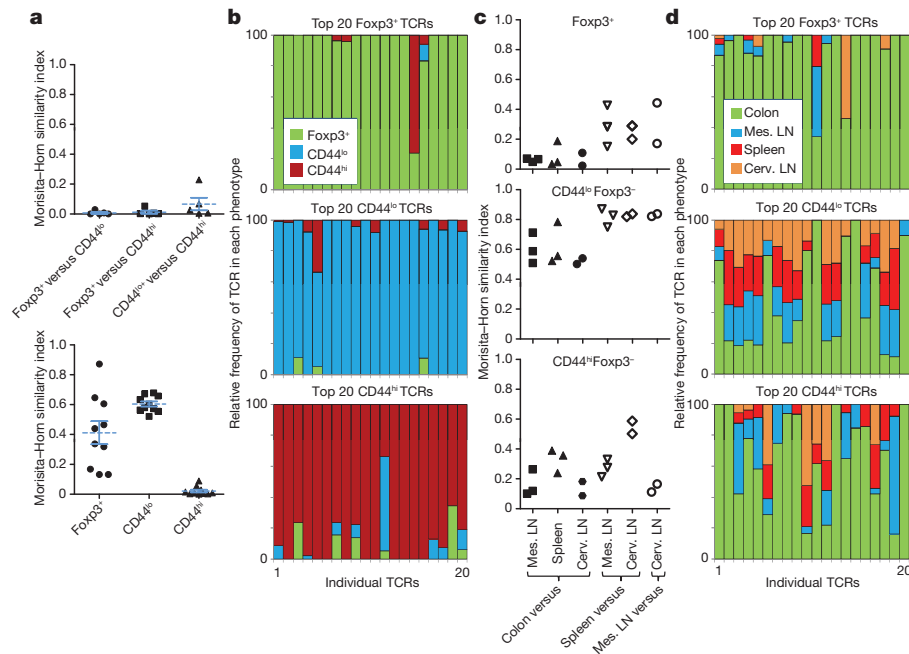


Figure 1 | The colonic T_{reg} TCR repertoire is unique. **a**, TCR usage between colonic T-cell subsets. A total of 2,892 TRAV14 TCR α sequences from colonic naive, memory and T_{reg} cells of five individual mice were compared (Supplementary Fig. 2). Each symbol represents a Morisita-Horn similarity comparison between two different T-cell subsets within each mouse (top), or a comparison of the same T-cell subset between different mice (bottom). Bars represent means \pm s.e.m. **b**, Analysis of individual TCRs. The relative distribution within all T-cell subsets is shown for the 20 most common individual TCRs in each colonic T-cell subset. For example, a TCR with equal percentage in the Foxp3⁺ and CD44^{hi} subset would be shown as a half-green/half-orange bar. This analysis uses the pooled data set, which includes

sequences from individual mice as well as 9,680 sequences from experiments 1–3, each consisting of cells from three to five mice (Supplementary Fig. 2). Note that one TCR is found in both Foxp3⁺ (no. 15) and CD44^{hi} (no. 3) plots, and one in both CD44^{lo} (no. 5) and CD44^{hi} (no. 12) plots; all others appear in only one plot. **c**, Anatomical distribution of colonic TCRs. Morisita-Horn indices comparing the colon data with those for other locations (filled symbols) or between each of the other locations (open symbols) are shown. Mes., mesenteric; cerv., cervical. **d**, Analysis of individual TCRs. The 20 most prevalent colon TCRs for each subset in the pooled data set are shown, and their presence at other locations is represented in a manner analogous to **b**. LN, lymph node.

recognized by CT7, but none of 34 other sequenced isolates, were identified as *Parabacteroides distasonis* (Fig. 2c and Supplementary Fig. 6). To assess whether CT7 broadly recognizes commensal species within the *Parabacteroides/Bacteroides* genera, we screened it against an additional panel of closely related mouse-derived commensal *Parabacteroides* and *Bacteroides* species²². CT7, but not CT2 or CT6, recognized only a subset of these bacterial species, including a second isolate of *P. distasonis* (Fig. 2d). Importantly, isolates that did not stimulate CT7 were recognized by another TCR hybridoma, DP1 (Supplementary Fig. 7a), indicating that these preparations contained antigens capable of stimulating TCRs. The almost mutually exclusive specificity of CT7 and DP1 within the Bacteroidaceae family makes it unlikely that these TCRs recognize self antigens in the host that are differentially induced by these closely related bacteria²² (Supplementary Fig. 7b). The TCR-specific reactivity patterns further suggest that TCR activation is not due to non-specific stimulation by generic immunostimulatory bacterial components or superantigens. Rather, these TCRs probably recognize distinct bacterial protein antigens, because predigestion of heat-killed *P. distasonis* with proteinase K abrogated recognition by the CT7 hybridoma (Supplementary Fig. 7c), in contrast with a report of T_{reg}-cell induction by protease-insensitive capsular polysaccharide from *B. fragilis*⁶. Although proof of direct bacterial recognition will require the identification of specific epitopes, these data strongly argue for the recognition of a bacteria-derived peptide by colonic T_{reg} TCR CT7.

More than half of the tested colonic T_{reg} TCRs recognized colonic contents and/or bacterial isolates (Table 1). However, this may underestimate the true frequency of colonic T_{reg} TCRs that respond to bacterial antigens. A lack of reactivity in our screen cannot be interpreted to mean that the TCR does not recognize bacteria, because the antigens may be rare in unfractionated colonic contents, lost on

autoclaving, or derived from an organism that was not isolated in our screen. Thus, the specificity of common colonic T_{reg} TCRs seems to be skewed towards the recognition of bacterial antigens.

Although T_{reg} cells may develop extrathymically as a result of an encounter with bacterial antigens, it is also possible that these T_{reg} cells are selected by self-antigen recognition in the thymus, followed by expansion in the periphery due to cross-reactivity. To assess the ability of these colonic T_{reg} TCRs to facilitate thymic T_{reg}-cell selection, we tracked the development of immature Foxp3^{gfp} Rag1^{-/-} thymocytes that were retrovirally transduced with a colonic T_{reg} TCR. None of the colonic T_{reg} TCRs generated an appreciable frequency of Foxp3⁺ thymocytes (Fig. 3a and Supplementary Figs 8 and 9a), in contrast with T_{reg} TCRs normally found at other peripheral locations (R19, G25 and R111; Supplementary Fig. 3). Note that G25 and R111 can be found at low frequency in the colonic T_{reg}-cell subset, suggesting that the colon does contain some thymically derived T_{reg} cells. The lack of T_{reg} development in cells expressing colonic T_{reg} TCRs cannot be attributed to an overwhelmed thymic niche^{23,24} (Supplementary Fig. 9b). These data therefore demonstrate that many common colonic T_{reg} TCRs facilitate thymic T_{reg}-cell selection poorly, if at all, implying that these TCRs instead mediate peripheral T_{reg}-cell development.

The retroviral transduction of thymocytes does not result in the emergence of sufficient numbers of transduced T cells from the thymus to permit their reliable detection in the periphery. We therefore retrovirally transduced self-renewing bone marrow progenitors and used them to create stable chimaeras, selecting colonic T_{reg} TCRs CT2 and CT6 on the basis of their *in vitro* reactivity to colonic contents (Fig. 2a). In these chimaeras we observed virtually no development of Foxp3 expression in cells expressing CT2 or CT6 (Fig. 3b, left; Supplementary Fig. 10). We reasoned that CT2 and CT6 may not recognize the microbiota in these commercially sourced host mice (Fig. 2a), and

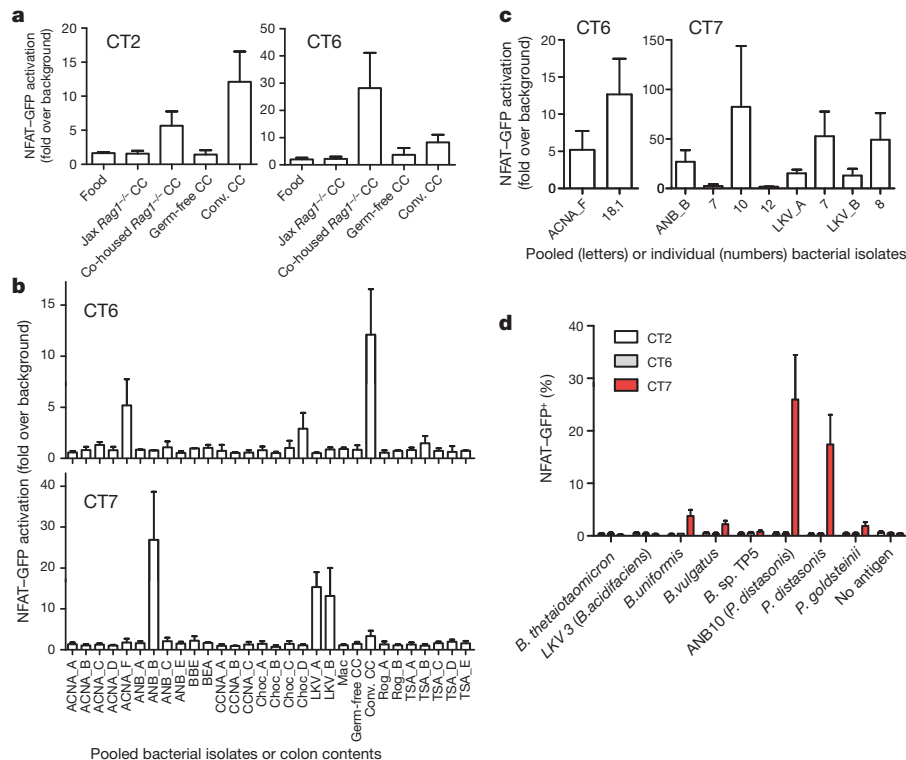


Figure 2 | *In vitro* reactivity of colonic T_{reg} TCRs to colonic contents and bacterial isolates. **a**, Reactivity to colonic contents. Colonic T_{reg} TCR-expressing NFAT-GFP hybridoma cells were cultured with Flt3L-induced dendritic cells in the presence of autoclaved food homogenate, or autoclaved colonic contents (CC) isolated from Rag1^{-/-} mice from Jackson Laboratories (Jax Rag1^{-/-}), Jax Rag1^{-/-} mice housed together with mice from our colony (co-housed Rag1^{-/-}), germ-free mice and conventionally housed (conv.) mice in our colony. **b**, Reactivity to bacterial pools. Cultures of heat-killed commensal bacteria isolated from our colony (Supplementary Fig. 6) were pooled (denoted by culture conditions and a letter) and screened for their

therefore performed experiments in which the chimaeras were housed together with mice from our colony. This resulted in the induction of CT2-expressing or CT6-expressing T_{reg} cells preferentially localized in the colon (Fig. 3b, right, and Supplementary Fig. 11). In an observation paralleling that previously made in the thymus^{23,24}, there also seems to be a saturable, antigen-specific T_{reg}-cell niche in the periphery²⁵ (Supplementary Fig. 11c).

Although these data strongly suggest that many colonic T_{reg} cells are generated extrathymically on encountering bacterial antigen, it remained possible that a rare population of thymically generated T_{reg} cells below the limit of our detection expanded on encountering peripheral antigen²⁵. We therefore generated CT6 TCR transgenic mice (Supplementary Fig. 12) and adoptively transferred CD44^{lo} Foxp3⁻ CT6 T cells mixed with congenic polyclonal CD4⁺ 'filler' T cells into T-cell-deficient *Tcrb*^{-/-} hosts. Consistent with the bone marrow chimaera data, the transferred CT6 T cells expanded and induced the

ability to stimulate colonic T_{reg} TCR-expressing hybridomas. For **a**, **b**, see Supplementary Fig. 4 for additional TCRs. **c**, Reactivity to individual isolates. Hybridomas showing reactivity against a pool of bacterial isolates were re-screened against the individual constituents (numbered). Data shown in **a**–**c** are the fold change in percentage GFP⁺ over the no-antigen control from 2–4 experiments (means ± s.e.m.). **d**, Specificity of colonic T_{reg} TCRs. A panel of heat-killed *Parabacteroides* and *Bacteroides* species (Supplementary Fig. 7) was tested against hybridomas expressing CT2, CT6 and CT7. Data shown are means ± s.e.m. for three experiments.

expression of Foxp3 only if the recipients were housed together with mice from our colony (Fig. 3c and Supplementary Fig. 13). Taken together with the observed lack of T_{reg}-cell development by thymocytes expressing colonic T_{reg} TCRs (Fig. 3a), these data suggest that a substantial proportion of the colonic T_{reg} population arises extrathymically from antigen-specific interactions with the colonic microbiota.

The notion that most colonic T_{reg} cells are generated as a result of microbial interactions is at odds with the observation that germ-free mice have normal T_{reg}-cell frequencies⁶ (Supplementary Fig. 1). However, we and others¹⁴ have observed that most colonic T_{reg} cells in conventionally housed, but not germ-free, mice are probably of peripheral origin, because these cells express low levels of the transcription factor Helios (Fig. 3d and Supplementary Fig. 14), a putative marker for thymically derived T_{reg} cells²⁶. We therefore speculate that germ-free conditions skew the colonic T_{reg} TCR repertoire towards thymically derived T_{reg} TCRs.

Table 1 | Summary of *in vitro* screening of colonic T_{reg}-cell TCRs

Name	CDR3 amino-acid residue sequence	Reactivity		
		Conv. CC	Transferred by co-housing	Bacterial isolate
CT1	AASWASGYNKLT	Yes	Yes	
CT2	AASAIWNTGYQNFY	Yes	Yes	
CT4	AASEYSALGRLH			
CT6	AASGYSALGRLH	Yes	Yes	<i>Clostridiales</i> sp. ACNA18.1
CT7	AASATGDNRF			<i>Parabacteroides distasonis</i>
CT8	AASLTGGYKVV			
CT9	AASADNRAGNKLT	Yes	Yes	
G57	AASELYQGGRALI	Yes		

Conv. CC, colonic contents from conventionally housed mice.

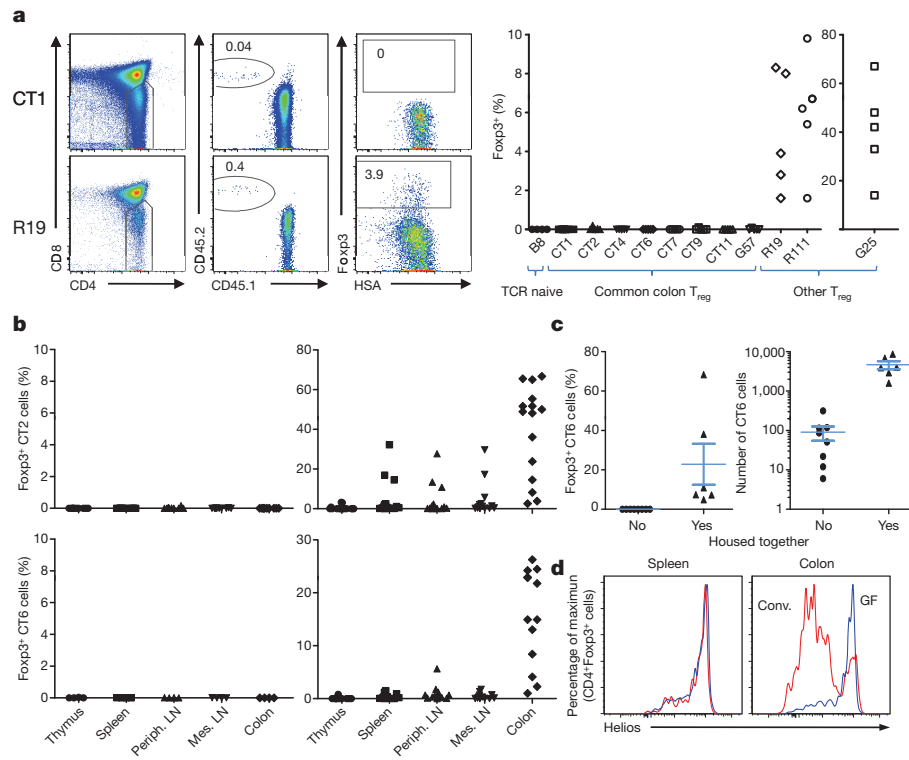


Figure 3 | Colonic T_{reg} TCRs facilitate thymic T_{reg}-cell development poorly, if at all. **a**, Assessment of thymic T_{reg}-cell development from TCR- $\alpha\beta$ -transduced *Rag1*^{-/-} thymocytes. The gating strategy (left), and summary of two to four experiments per TCR (right) are shown. See Supplementary Fig. 3 for additional TCR information, and Supplementary Figs 8 and 9 for plots and analysis of clonal frequencies. Comparison of colon versus other T_{reg} TCRs revealed *P*-values < 0.01. **b**, Mixed retroviral bone-marrow chimaeras. The percentage of Foxp3⁺ cells in the CT2-expressing or CT6-expressing CD45.2⁺CD4⁺ population is shown in hosts housed together (right) or not (left) with mice from our colony. See Supplementary Figs 10 and 11 for additional analyses. Periph., peripheral. **c**, Peripheral conversion of CT6 TCR

transgenic cells. Naive CD45.2 CT6 cells and CD45.1 CD4⁺ filler cells were adoptively transferred into *Tcrb*^{-/-} hosts for 5 weeks. The percentage (left) of Foxp3⁺ CT6 cells (V α 2⁺V β 6⁺CD45.2⁺CD45.1⁻CD4⁺) are shown. The number of CT6 cells (right) was determined by flow cytometry of the entire colonic lamina propria. Data are from three experiments; bars represent means \pm s.e.m. See Supplementary Fig. 13 for flow cytometric plots. **d**, Helios expression in T_{reg} cells. Representative intracellular Helios staining in CD4⁺Foxp3⁺ cells from conventionally housed (conv.) and germ-free (GF) *Foxp3*^{gfp} mice is shown, and is summarized in Supplementary Fig. 14. For all plots in this figure, each symbol represents data from an individual host.

The efficient differentiation of naive T cells into T_{reg} rather than effector, cells may be important for generating colonic tolerance, because it has been observed that TCRs that facilitate thymic T_{reg}-cell development can be pathogenic when expressed on effector T cells^{17,27}. To address this possibility, we performed an initial analysis of colonic TCR repertoires in mice expressing the fixed TCR β chain and undergoing spontaneous colitis due to genetic deficiencies in interleukin (IL)-2, IL-10 or transforming growth factor- β receptor signalling (Supplementary Fig. 2a). We observed that several colonic TCRs almost exclusively found in the Foxp3⁺ data sets in normal mice were found in the effector/memory data sets in the diseased animals (Supplementary Fig. 15a). Although these genetic manipulations may affect T_{reg}-cell development or survival, the relatively high abundance of some of these TCRs in the CD44^{hi} subset suggests that effector cells expressing these TCRs are expanding in the colitic environment. To test for the pathogenic potential of colonic T_{reg} TCRs, we retrovirally expressed CT2 and CT6 TCRs on peripheral, monospecific TCR $\alpha\beta$ transgenic cells with known specificity for a foreign antigen (human CLIP peptide). Adoptive transfer of these cells, which were virtually all Foxp3⁻, into *Rag1*^{-/-} hosts housed with mice from our colony induced weight loss and colitis (Fig. 4 and Supplementary Fig. 15b, c). In contrast, cells expressing only the transgenic TCR or a TCR from the naive T-cell subset (B8) did not. The failure of these retrovirally transduced T cells to upregulate Foxp3 and become regulatory in this situation is probably due to expansion in a lymphopenic environment, as well as *in vitro* T-cell activation—a requirement for retroviral

transduction. Thus, these data illustrate the potential pathological consequences of T-cell recognition of commensal bacterial antigens under conditions that disfavour T_{reg} development.

This analysis of common colonic T_{reg} TCRs in a fixed TCR β repertoire suggests a model (Supplementary Fig. 16) in which T cells expressing these TCRs exist as naive T cells in the absence of antigen

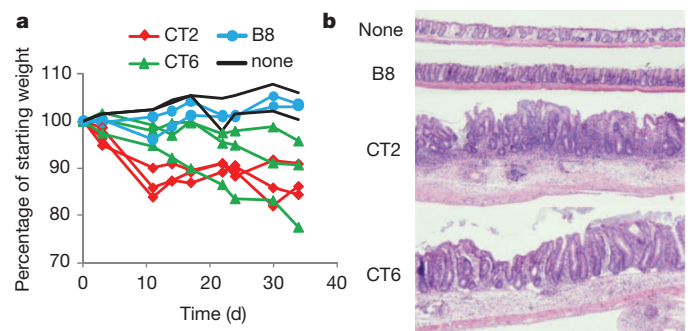


Figure 4 | Pathogenic potential of colonic T_{reg} TCRs. **a**, Adoptive transfer of peripheral T cells transduced with CT2 or CT6 into *Rag1*^{-/-} hosts housed with mice from our colony. Non-transduced T cells (none) or naive TCR transduced T cells (B8) were used as controls. Each line represents an individual recipient. One representative experiment is shown (summary in Supplementary Fig. 15b). **b**, A representative haematoxylin/eosin-stained section of the descending colon 7–10 weeks after T-cell transfer. Original magnification $\times 4$.

(Fig. 3b, left, and Supplementary Figs 10 and 11). Encounter with bacteria-derived foreign antigens in the colon seems to drive the generation of Foxp3⁺ T_{reg} cells efficiently (Fig. 3b), because it typically does not result in the substantial simultaneous development of CD44^{hi} cells of the same specificity (Fig. 1a, b). This diversion of naive T cells with bacterial TCR specificity into the T_{reg}-cell lineage may be crucial for preventing the generation of colitogenic effector cells (Fig. 4). Thus, these data support a model in which an individual's T-cell population is not only instructed by classic self/non-self discrimination mechanisms during thymic development but is also educated in the periphery to accommodate the variety of non-self antigens derived from the commensal microbiota at mucosal sites.

METHODS SUMMARY

Mice. TCRβ Foxp3^{gfp} Tcrα^{+/-}; Foxp3^{IRES-GFP}; Il-2^{-/-}; Il-10^{-/-}; and dnTGFβRII strains have been described (see Methods). C57BL/6 Rag1^{-/-} and CD45.1 mice were obtained from Jackson Laboratories and the National Cancer Institute, respectively. Germ-free mice were generated in collaboration with J. Gordon. CT6 transgenic mice were generated as described²³.

TCR repertoire analysis. Analyses of TRAV14 (Vα2) TCR sequences from TCRβ transgenic mice were performed as described²⁰. Lamina propria cell suspensions from the entire colon were prepared as described⁹, and CD4⁺ subsets were sorted with a FACSAria (Becton Dickinson).

Hybridoma assays. Hybridoma cells expressing GFP under an NFAT promoter²¹ were retrovirally transduced with TCRβ-IRES-mCD4 and an individual TCR α chain. Hybridomas were cultured with flt3-ligand-elicited dendritic cells with the indicated antigen preparations and analysed by flow cytometry after 1.5 days.

Antigen preparations. Whole colonic contents and food pellets were diluted with phosphate-buffered saline, vortex-mixed, homogenized, filtered, and autoclaved for 15 min. Colonic bacterial isolation was performed as described²² (see Supplementary Fig. 6).

In vivo developmental assays. Retroviral transduction and intrathymic transfer of Rag1^{-/-} thymocytes were performed as described in Methods. Analysis of CD4 SP thymocytes was performed about 2.5 weeks later. Retroviral bone marrow chimaeras were created as described²³. Some recipients were housed together with mice from our colony 2 weeks after bone marrow reconstitution, for a period of 1 week.

In vivo peripheral T-cell assays. Retroviral transduction of peripheral TCRβ-Foxp3^{gfp} Rag1^{-/-} T cells was performed as described²⁷, and cells were transferred intravenously into Rag1^{-/-} hosts housed with them. Sorted CD4⁺CD44^{lo}Foxp3⁻ cells (5 × 10⁴) from CT6 transgenic mice were co-transferred with 5 × 10⁵ CD45.1⁺CD4⁺ 'filler' cells into Tcrb^{-/-} mice. Recovered cells were analysed by flow cytometry 5 weeks later.

Statistics. The Wilcoxon rank sum test was used unless otherwise indicated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 November 2010; accepted 5 August 2011.

Published online 21 September 2011.

1. Josefowicz, S. Z. & Rudensky, A. Control of regulatory T cell lineage commitment and maintenance. *Immunity* **30**, 616–625 (2009).
2. Belkaid, Y. & Tarbell, K. Regulatory T cells in the control of host–microorganism interactions. *Annu. Rev. Immunol.* **27**, 551–589 (2009).
3. Barnes, M. J. & Powrie, F. Regulatory T cells reinforce intestinal homeostasis. *Immunity* **31**, 401–411 (2009).
4. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host–bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
5. Min, B. *et al.* Gut flora antigens are not important in the maintenance of regulatory T cell heterogeneity and homeostasis. *Eur. J. Immunol.* **37**, 1916–1923 (2007).
6. Round, J. L. & Mazmanian, S. K. Inducible Foxp3⁺ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **107**, 12204–12209 (2010).

7. Singh, B. *et al.* Control of intestinal inflammation by regulatory T cells. *Immunol. Rev.* **182**, 190–200 (2001).
8. Curotto de Lafaille, M. A. *et al.* Adaptive Foxp3⁺ regulatory T cell-dependent and -independent control of allergic inflammation. *Immunity* **29**, 114–126 (2008).
9. Sun, C. M. *et al.* Small intestine lamina propria dendritic cells promote *de novo* generation of Foxp3 Treg cells via retinoic acid. *J. Exp. Med.* **204**, 1775–1785 (2007).
10. Zheng, Y. *et al.* Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. *Nature* **463**, 808–812 (2010).
11. Coombes, J. L. *et al.* A functionally specialized population of mucosal CD103⁺ DCs induces Foxp3⁺ regulatory T cells via a TGF-β and retinoic acid-dependent mechanism. *J. Exp. Med.* **204**, 1757–1764 (2007).
12. Mucida, D. *et al.* Reciprocal T_H17 and regulatory T cell differentiation mediated by retinoic acid. *Science* **317**, 256–260 (2007).
13. Benson, M. J., Pino-Lagos, K., Roseblatt, M. & Noelle, R. J. All-trans retinoic acid mediates enhanced Treg cell growth, differentiation, and gut homing in the face of high levels of co-stimulation. *J. Exp. Med.* **204**, 1765–1774 (2007).
14. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
15. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332**, 974–977 (2011).
16. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell–IgA response to the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **106**, 19256–19261 (2009).
17. Hsieh, C.-S. *et al.* Recognition of the peripheral self by naturally arising CD25⁺CD4⁺ T cell receptors. *Immunity* **21**, 267–277 (2004).
18. Pacholczyk, R., Ignatowicz, H., Kraj, P. & Ignatowicz, L. Origin and T cell receptor diversity of Foxp3⁺CD4⁺CD25⁺ T cells. *Immunity* **25**, 249–259 (2006).
19. Wong, J. *et al.* Adaptation of TCR repertoires to self-peptides in regulatory and nonregulatory CD4⁺ T cells. *J. Immunol.* **178**, 7032–7041 (2007).
20. Lathrop, S. K., Santacruz, N. A., Pham, D., Luo, J. & Hsieh, C. S. Antigen-specific peripheral shaping of the natural regulatory T cell population. *J. Exp. Med.* **205**, 3105–3117 (2008).
21. Ise, W. *et al.* CTLA-4 suppresses the pathogenicity of self antigen-specific T cells by cell-intrinsic and cell-extrinsic mechanisms. *Nature Immunol.* **11**, 129–135 (2010).
22. Bloom, S. M. *et al.* Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**, 390–403 (2011).
23. Bautista, J. L. *et al.* Intracolon competition limits the fate determination of regulatory T cells in the thymus. *Nature Immunol.* **10**, 610–617 (2009).
24. Leung, M. W., Shen, S. & Lafaille, J. J. TCR-dependent differentiation of thymic Foxp3⁺ cells is limited to small clonal sizes. *J. Exp. Med.* **206**, 2121–2130 (2009).
25. Nishio, J., Feuerer, M., Wong, J., Mathis, D. & Benoist, C. Anti-CD3 therapy permits regulatory T cells to surmount T cell receptor-specified peripheral niche constraints. *J. Exp. Med.* **207**, 1879–1889 (2010).
26. Thornton, A. M. *et al.* Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3⁺ T regulatory cells. *J. Immunol.* **184**, 3433–3441 (2010).
27. Hsieh, C. S., Zheng, Y., Liang, Y., Fontenot, J. D. & Rudensky, A. Y. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nature Immunol.* **7**, 401–410 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Murphy, T. Egawa, Y. Zheng, J. Scott-Browne, J. Fontenot and S. Wetzel for discussion and reading of the manuscript; A. Kau and J. Gordon for discussions and generation of germ-free animals; N. P. Malvin for assistance with bacteriology; and J. Hunn for technical assistance. C.S.H. and co-workers are funded by the National Institute of Allergy and Infectious Diseases and the Burroughs-Wellcome Fund. S.M.B. was supported by National Institutes of Health training grant 5T32AI0071632.

Author Contributions S.K.L., S.R., K.N. and N.S. performed most of the experiments. S.M.B. designed and performed the bacteriology. C.W.L. developed and assisted with the intrathymic transfer experiments. D.P. and T.S. were involved in study design. S.K.L. and C.S.H. designed the experiments and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.S.H. (chsieh@wustl.edu).

METHODS

Mice. TCR β *Foxp3*^{gfp} *Tcr* α ^{+/-} (ref. 27), *Foxp3*^{IRES-GFP} (ref. 28), *Il-2*^{-/-} (ref. 29), *Il-10*^{-/-} (ref. 30) and dnTGF β RII (ref. 31) strains on C57BL/6 background have been described. C57BL/6 *Rag1*^{-/-} and C57BL/6 CD45.1 congenic mice were obtained from Jackson Laboratories and National Cancer Institute, respectively. Germ-free mice were generated in collaboration with J. Gordon. CT6 transgenic mice were generated as described²³, with microinjection into B6 \times 129 fertilized eggs. In experiments in which animals obtained from commercial vendors were housed together, a female bred in our facility was added to each cage of mice for 7 days, and then removed before the experiment. In retroviral bone marrow chimaeras, the animal was added 2 weeks after bone marrow transfer, for 1 week, then removed for the duration of the experiment. Animal experiments were performed in a specific pathogen-free facility in accordance with the guidelines of the Institutional Animal Care and Use Committee at Washington University.

TCR repertoire analysis. Analyses of TRAV14 (V α 2) TCR sequences from TCR β transgenic mice were performed as described²⁰. In brief, lamina propria cell suspensions were prepared from the entire colon (including the caecum) as described⁹, without the final Percoll enrichment, and sorted into CD4⁺ T-cell subsets (CD44^{hi}Foxp3⁻, CD44^{lo}Foxp3⁻ and Foxp3⁺) with a FACSARIA (Becton Dickinson). RNA was isolated, TCR α cDNA was isolated, and a TRAV14 (V α 2) cDNA library was generated by PCR and sequenced by the Genome Sequencing Center at Washington University. Comparison of TCR repertoires was performed with the Morisita–Horn statistical test³², which compares two populations, taking into account the overlap and relative abundance of the species in the two populations, and expresses their similarity on a scale from 0 (no similarity) to 1.0 (exactly the same).

Hybridoma assays. T-cell hybridoma cells, which do not express a T-cell receptor and express GFP under an NFAT promoter²¹, were retrovirally transduced with TCR β -IRES-mCD4 and sorted for CD4 expression. TCR α chains of interest, selected on the basis of average frequency and presence in at least two independent data sets, were then retrovirally introduced, and V α 2-expressing cells were sorted. The hybridomas differ only in their retrovirally encoded TCR α chain, which does not confer any functional properties of the cell from which the TCR originated, thereby allowing a direct comparison of antigen recognition between TCRs from different T-cell subsets. These hybridoma cells ((2.5–5.0) \times 10⁴) were cultured with flt3-ligand-elicited dendritic cells (5 \times 10⁴) and the indicated antigen preparations, diluted 1:200, in flat-bottomed 96-well plates. The V α 2⁺CD4⁺ cells were analysed for GFP expression after 1.5 days by flow cytometry on a FACSCanto or FACSARIA (Becton Dickinson). Controls without antigen were performed in parallel and used to normalize data against variations in background fluorescence between experiments.

Antigen preparations. Whole colonic contents were scraped from longitudinally opened colon and caecum; they were diluted with phosphate-buffered saline, then vortex-mixed, homogenized, filtered through a 70- μ m mesh, and autoclaved for 15 min. Food antigen was also diluted with phosphate-buffered saline, homogenized, filtered and autoclaved. Preparations were stored short-term at -20 °C and long-term at -80 °C. For a description of the colonic bacterial isolation see Supplementary Fig. 6 and ref. 22. Isolates were named according to their culture conditions followed by a number; pools of two or three isolates were designated by a letter.

Thymic T_{reg}-cell developmental assay. Retroviral transduction of *Foxp3*^{gfp}*Rag1*^{-/-} thymocytes was performed as described³³ by using a TCR α -P2A-TCR β vector³⁴. Thymocytes were injected into the thymus of sublethally irradiated (600 rad) CD45.1 congenic recipients; 14–16 days later, CD45.2⁺CD4⁺CD8⁻ thymocytes were analysed by flow cytometry for expression of *Foxp3*^{gfp} with the use of a FACSARIA. About half of each thymus (8 \times 10⁷) was analysed by flow cytometry.

A total of 10⁶ events were initially collected to determine the frequency of CD45.1 and CD45.2 cells. Subsequently, the storage gate was changed to include only CD45.2 cells for the rest of the sample at high speed (about 25,000–30,000 events s⁻¹). Of the 45 recipients of cells expressing colonic T_{reg} TCRs, 4 of 45 showed a handful of Foxp3⁺ events (1/5 G57, 1/5 CT9, 2/5 CT2). We believe that many or all of these may be artefacts, because the Foxp3⁺ events shown in Supplementary Fig. 8b have a larger than usual Forward Scatter (FSC) and Side Scatter (SSC), nor can they be found in ten CT2 retroviral bone marrow chimaeras (Supplementary Fig. 10, top). However, we cannot be certain that these events are artefacts, because they are few and not reproducibly observed in all recipients. The sensitivity for picking up Foxp3⁺ cells may be estimated from the binomial distribution, in which a population frequency of 0.3% would result in at least 1 event per 1,000 with 95% confidence, and can be calculated from the number of CD4SP events processed (Supplementary Fig. 9b, bottom). However, increasing the number of CD4SP cells expressing a particular TCR may not increase the sensitivity of the assay, because it seems that T_{reg} TCRs often show an inverse correlation between clonal frequency and T_{reg}-cell development^{23,24}.

Retroviral bone marrow chimaeras. *Foxp3*^{IRES-GFP} *Rag1*^{-/-} (CD45.2) bone marrow was retrovirally transduced with the CT2 or CT6 TCR using a TCR α -P2A-TCR β vector³⁴ and injected intravenously with wild-type CD45.1 bone marrow into lethally irradiated (1,050 rad) CD45.1 hosts from NCI to create chimaeras as described²³. Some recipients were housed together (co-housed) with mice from our colony for a period of 1 week, beginning 2 weeks after bone marrow reconstitution. Cells isolated from the thymus, spleen, mesenteric lymph node and colon were analysed by flow cytometry after 6–8 weeks.

In vivo TCR-induced pathology. Retroviral transduction of *in vitro* activated peripheral TCR α - β *Foxp3*^{gfp} *Rag1*^{-/-} T cells was performed as described²⁷. The entire T-cell population was transferred, and adjusted such that 6 \times 10⁵ transduced cells (14% mean transduction efficiency, range 5.8–22%) were intravenously transferred into each co-housed *Rag1*^{-/-} mouse. After 7–10 weeks, tissue from the caecum and colon underwent fixation and staining with haematoxylin and eosin.

CT6 TCR transgenic cell transfer experiment. Naive (CD4⁺CD44^{lo}Foxp3⁻) T cells were sorted from lymph node and spleen from CT6 *Foxp3*^{IRES-GFP} *Rag1*^{-/-} mice. Cells (5 \times 10⁴) were transferred simultaneously with 5 \times 10⁵ CD45.1⁺CD4⁺ congenic ‘filler’ cells into *Tcrb*^{-/-} mice, with the notion that providing filler cells, which include T_{reg} cells, may limit lymphopenic expansion and facilitate conversion. V β 6⁺V α 2⁺CD45.2⁺CD45.1⁻CD4⁺ CT6 Tg T cells in the colonic lamina propria were assessed for Foxp3 expression 5 weeks later.

Statistics. The Wilcoxon rank sum test was used unless otherwise indicated.

28. Lin, W. et al. Regulatory T cell development in the absence of functional Foxp3. *Nature Immunol.* **8**, 359–368 (2007).
29. Schorle, H., Holtschke, T., Hunig, T., Schimpl, A. & Horak, I. Development and function of T cells in mice rendered interleukin-2 deficient by gene targeting. *Nature* **352**, 621–624 (1991).
30. Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K. & Muller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
31. Gorelik, L. & Flavell, R. A. Abrogation of TGF β signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* **12**, 171–181 (2000).
32. Magurran, A. E. *Ecological Diversity and its Measurement* (Princeton Univ. Press, 1988).
33. Haxhinasto, S., Mathis, D. & Benoist, C. The AKT-mTOR axis regulates de novo differentiation of CD4⁺ Foxp3⁺ cells. *J. Exp. Med.* **205**, 565–574 (2008).
34. Holst, J., Vignali, K. M., Burton, A. R. & Vignali, D. A. Rapid analysis of T-cell selection *in vivo* using T cell-receptor retrogenic mice. *Nature Methods* **3**, 191–197 (2006).

A reserve stem cell population in small intestine renders *Lgr5*-positive cells dispensable

Hua Tian¹, Brian Biehs², Søren Warming¹, Kevin G. Leong³, Linda Rangell⁴, Ophir D. Klein² & Frederic J. de Sauvage¹

The small intestine epithelium renews every 2 to 5 days, making it one of the most regenerative mammalian tissues. Genetic inducible fate mapping studies have identified two principal epithelial stem cell pools in this tissue. One pool consists of columnar *Lgr5*-expressing cells that cycle rapidly and are present predominantly at the crypt base¹. The other pool consists of *Bmi1*-expressing cells that largely reside above the crypt base². However, the relative functions of these two pools and their interrelationship are not understood. Here we specifically ablated *Lgr5*-expressing cells in mice using a human diphtheria toxin receptor (*DTR*) gene knocked into the *Lgr5* locus. We found that complete loss of the *Lgr5*-expressing cells did not perturb homeostasis of the epithelium, indicating that other cell types can compensate for the elimination of this population. After ablation of *Lgr5*-expressing cells, progeny production by *Bmi1*-expressing cells increased, indicating that *Bmi1*-expressing stem cells compensate for the loss of *Lgr5*-expressing cells. Indeed, lineage tracing showed that *Bmi1*-expressing cells gave rise to *Lgr5*-expressing cells, pointing to a hierarchy of stem cells in the intestinal epithelium. Our results demonstrate that *Lgr5*-expressing cells are dispensable for normal intestinal homeostasis, and that in the absence of these cells, *Bmi1*-expressing cells can serve as an alternative stem cell pool. These data provide the first experimental evidence for the interrelationship between these populations. The *Bmi1*-expressing stem cells may represent both a reserve stem cell pool in case of injury to the small intestine epithelium and a source for replenishment of the *Lgr5*-expressing cells under non-pathological conditions.

Two types of stem cells have been described in the small intestine based on location and cycling dynamics^{1–4}. Fast-cycling stem cells express markers including *Lgr5*, *Cd133* (also known as *Prom1*) and *Sox9* (refs 1, 5, 6) and are present throughout the intestine. Also known as crypt base columnar cells (CBCs), these slender cells populate the crypt and villi within 3 days, and are interspersed among the Paneth cells that support them^{7,8}. Slower-cycling stem cells, marked by enriched expression of *Bmi1* or mouse *Tert* (*mTert*), represent a rarer cell population^{2,9}. These cells form a descending gradient from proximal to distal regions of the intestine, such that they are more prevalent in the duodenum than in the colon. Despite their rarity, *Bmi1*-expressing stem cells are crucial for crypt maintenance².

To study the function of *Lgr5*-expressing cells, we replaced the first coding exon of *Lgr5* with two distinct cassettes. The first consisted of a dsRED-IRES-CreERT2 sequence to enable genetic lineage tracing studies by tamoxifen (TAM)-inducible expression of Cre in *Lgr5*-expressing cells (Supplementary Fig. 1a, *Lgr5*^{CreER} allele). The second cassette contained enhanced green fluorescent protein (EGFP) linked in frame to a human DTR cDNA (Supplementary Fig. 1a, *Lgr5*^{DTR} allele), producing a fusion protein. Consistent with previous reports¹, one injection of TAM in *Lgr5*^{CreER};R26R mice marked *Lgr5*-expressing stem cells in a mosaic fashion and led to generation of labelled progeny for more than 60 days (Supplementary Fig. 1b). Expression of DTR-EGFP

in *Lgr5*^{DTR} mice functioned as a reporter for *Lgr5* expression (Fig. 1a) and also conferred diphtheria toxin (DT) sensitivity on CBCs. Expression of EGFP in mice carrying the *Lgr5*^{DTR} allele was detected at the membrane of cycling CBCs in every crypt (Supplementary Fig. 1c–e, CBCs are marked by asterisks).

We next set out to test the effects of eliminating *Lgr5*-expressing cells by administering DT to *Lgr5*^{DTR} mice. Twenty-four hours after DT administration, all EGFP-positive cells were depleted, including CBCs (Fig. 1a, b, j, k, p, q). Loss of *Lgr5*-expressing cells was further confirmed by the absence of *Lgr5* messenger RNA (Fig. 1d, e) and was accompanied by extensive apoptosis at the base of the crypts, with shedding of dead cells into the lumen (Fig. 1m, n).

After sustained DT exposure for 10 days, both the EGFP reporter and *Lgr5* mRNA were completely absent from the base of the crypts (Fig. 1c, f and Supplementary Fig. 2) but, notably, crypt architecture was comparable to controls (Fig. 1g, i, j, l). Proliferating CBCs were absent from the crypt (Fig. 1l, r), such that the crypt base was occupied mostly or entirely by Paneth cells (Supplementary Fig. 3a, b). The extensive apoptosis detected 24 h after DT treatment had significantly decreased by day 10 (compare Fig. 1n with o) but was still detectable. No increase in crypt fission after DT treatment was observed by haematoxylin and eosin staining at any time point (Fig. 1g–i).

Because *Lgr5*-expressing cells have been proposed to have a critical role in renewal of the intestine, it was surprising that the architecture of the intestinal epithelium was essentially intact after ablation of *Lgr5*-expressing CBCs (Fig. 1g–i). Within the villi, very little change in the total number of endocrine cells was observed (Supplementary Fig. 3c, d), and goblet cells were abundant in the crypts and villi (Supplementary Fig. 3g, h, j). Upon CBC ablation, Paneth cells were found at the bottom of the crypts and in some cases were mislocalized to the villi (Supplementary Fig. 3a, b and data not shown); additionally, migration of cells as assessed by BrdU pulse-chase labelling was normal (Supplementary Fig. 4). The only major difference from controls that we observed was in the secretory cell lineage; the number of chromogranin-A-positive enteroendocrine cells in the crypts doubled after DT administration for 10 days (Supplementary Fig. 3e, f, i).

We did not detect any *Lgr5*-expressing CBCs using either the EGFP reporter or *in situ* hybridization after 10 days of DT (Fig. 1c, f and Supplementary Fig. 2), but it was still possible that a few CBCs could have escaped ablation and repopulated the epithelium, as a similar scenario was reported in *c-Myc* and *Ascl2* conditional null mice^{10,11}. To address this possibility directly, we visualized *Lgr5*-expressing cell activity during DT selection by producing *Lgr5*^{DTR/CreER};R26R mice. These mutant mice carried two null alleles at the *Lgr5* locus, of which one enabled ablation of *Lgr5*-expressing cells and the other enabled lineage tracing of any possibly remaining *Lgr5*-expressing cells. These mice died at postnatal day (P)1, consistent with previous reports that *Lgr5* null mice are not viable¹². To analyse the postnatal gut, we grew pieces of small intestine from embryonic day (E)15 *Lgr5*^{DTR/CreER};R26R embryos under the kidney capsule of immunocompromised mice for

¹Department of Molecular Biology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ²Departments of Orofacial Sciences and Pediatrics, Institute for Human Genetics and Program in Craniofacial and Mesenchymal Biology, UCSF, 513 Parnassus Avenue, San Francisco, California 94143-0442, USA. ³Department of Research Oncology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ⁴Department of Pathology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA.

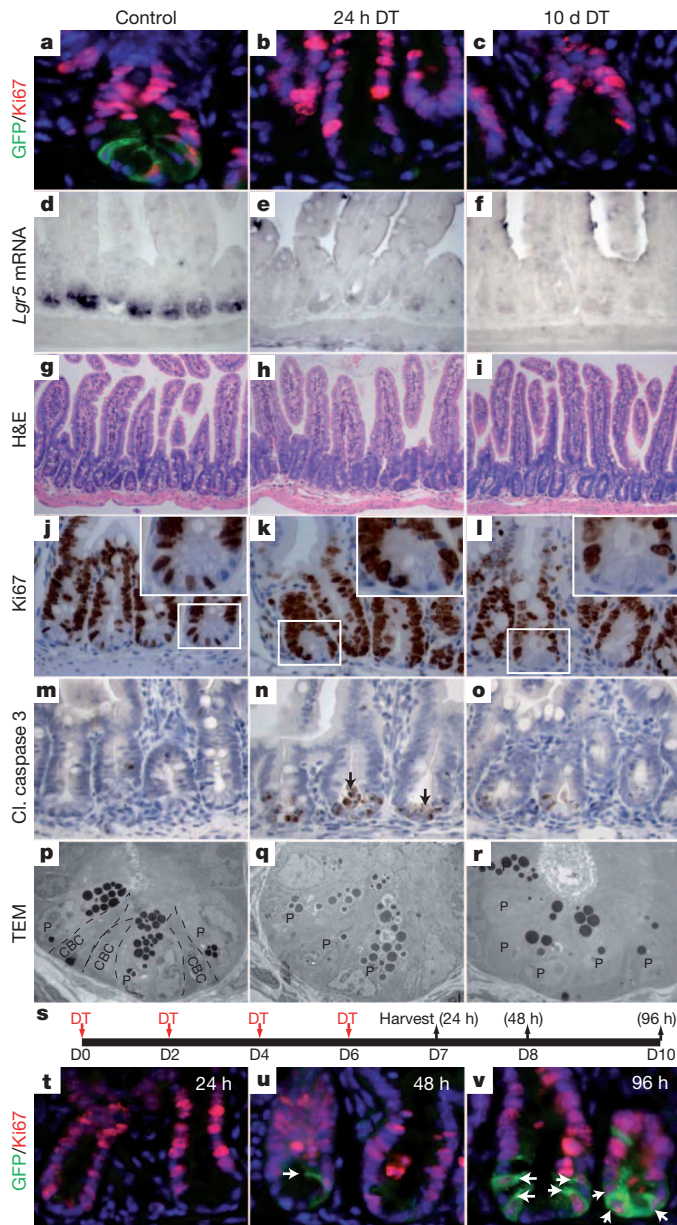


Figure 1 | Characterization of DT-mediated CBC ablation. **a**, EGFP is detected on the membrane of Ki67⁺ proliferating CBCs in saline-treated *Lgr5*^{DTR/+} mice. **b**, One dose of DT eliminates all DTR-EGFP-positive cells at 24 h. **c**, DT treatment for up to 10 days prevents reappearance of *Lgr5*-expressing cells. **d–f**, *Lgr5* mRNA is normally present at the bottom of the crypts (**d**) and is not detected after 24 h (**e**) or 10 days of DT treatment (**f**). **g–i**, Crypt architecture is intact after ablation of *Lgr5*-expressing CBCs. H&E, haematoxylin and eosin. **j–l**, Proliferation above the crypt base is normal after ablation of *Lgr5*-expressing CBCs. **m–o**, Extensive apoptosis is observed at the crypt base 24 h after DT and tapers off by 10 days, but is still higher than controls. 'Cl. caspase 3' is cleaved caspase 3. **p–r**, Electron microscopy shows that CBCs in controls are characterized by slender nuclei and scant cytoplasm. No CBCs remain at the crypt base after one dose or 10 days of DT treatment. The crypt base is filled with granule-rich Paneth cells. TEM, transmission electron microscopy. **s**, Dosing regimen for study of the recovery of *Lgr5*-expressing CBCs. **t**, No CBCs are detected 24 h after DT administration. **u**, A few *Lgr5*⁺/Ki67⁺ CBCs (arrow) were detected 48 h after the last dose of DT. **v**, More *Lgr5*⁺/Ki67⁺ CBCs (arrows) recovered after 96 h. Original magnification for panels: **a–f**, **m–o** and **t–v** at 40×; **g–i** at 20×; **j–l** at 63×; and **p–r** at 2,650×.

three weeks, at which point they formed crypts comparable to P17 intestine (Fig. 2a–e)¹³. After 10 days of TAM treatment, columns of blue cells emanated from the crypt base, and progeny of *Lgr5*-expressing cells differentiated into all four major cell types of the intestinal

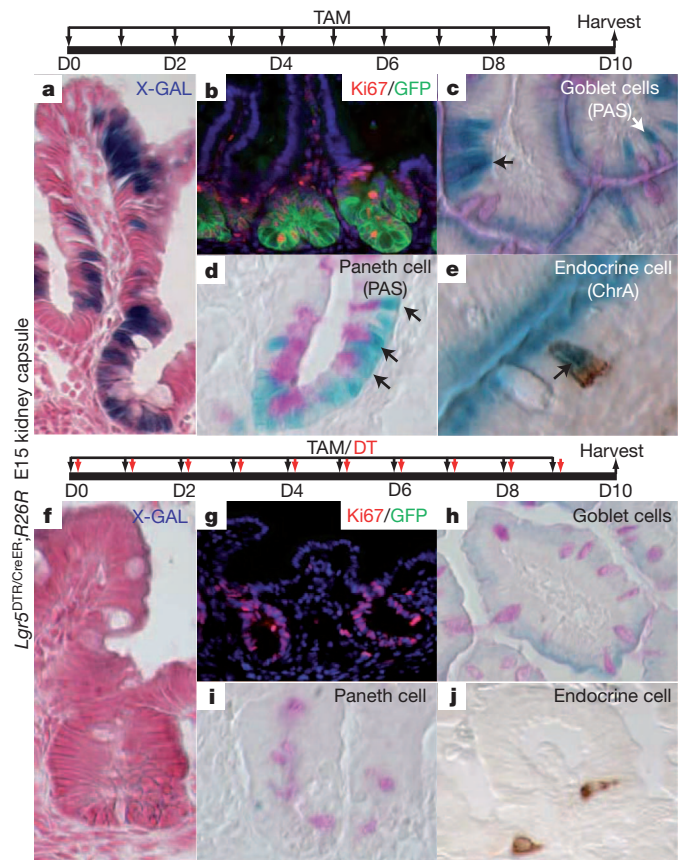


Figure 2 | Maintenance of normal crypt architecture is not mediated by *Lgr5*-positive cells that have escaped ablation. **a**, Ten-day lineage tracing of descendants of *Lgr5*-expressing stem cells shows a blue ribbon emanating from the base of the crypt in a grafted intestine piece from E15 *Lgr5*^{DTR/CreER} embryos. **b–e**, Normal proliferation and differentiation of intestinal epithelium after loss of *Lgr5* gene function. *Lgr5*-expressing stem cells can give rise to all four major differentiated cell types (arrows). X-GAL-positive cells mark *Lgr5*-positive stem cell progeny, which overlap with differentiated cell markers for goblet (**c**), Paneth (**d**) and endocrine cell (**e**) lineages. PAS, periodic acid Schiff; ChrA, Chromogranin A. **f**, Concurrent TAM and DT treatment kills all *Lgr5*-expressing cells. No progeny of *Lgr5*-expressing cells (blue) are detected in the grafted intestine. **g–j**, No GFP-positive cells are detected but proliferation and differentiation are normal after DT-mediated ablation of *Lgr5*-expressing CBCs. Original magnification for panels: **a**, **b**, **f**, **g** at 40×; **c–e**, **h–j** at 63×.

epithelium (Fig. 2a–e). Concomitant administration of DT and TAM for 10 days eradicated all EGFP-positive CBCs (Fig. 2g), and no cells descended from *Lgr5*-expressing cells were observed (Fig. 2f), confirming that the *Lgr5*^{DTR} allele leads to complete elimination of these cells. Importantly, no abnormalities in graft morphology, differentiation or proliferation were observed in these mice compared to controls (Fig. 2a–j).

Although *Lgr5*-expressing cells were completely depleted within 24 h of DT treatment, persistence of apoptotic bodies at the crypt base throughout the 10-day DT treatment suggested that *Lgr5*-expressing CBCs were continuously generated and eliminated during the treatment (Fig. 1n, o). This notion was supported by the quick recovery of *Lgr5*-expressing cells between 48 to 96 h after the final dose of DT (Fig. 1s–v). To follow the fate of the newly generated *Lgr5*-expressing cells, mice implanted with *Lgr5*^{DTR/CreER};R26R embryonic intestine fragments in the kidney capsule were allowed to recover for 6 days in the presence of TAM following 6 days of DT treatment. A row of blue cells emanated from the crypt base (Supplementary Fig. 5a), indicating that the newly formed *Lgr5*-expressing stem cells (Supplementary Fig. 5b, GFP-positive cells) gave rise to progeny that migrated out of the crypt. When the converse experiment was performed by

injecting TAM for 6 days and then dosing with DT from days 6 to 12, blue cells were only present in the upper region of the villi (Supplementary Fig. 5c), indicating that progeny of *Lgr5*-expressing cells marked between day 1 and 6 migrated out of the crypts into the villi, but that during DT treatment between days 6 and 12, *Lgr5*-expressing stem cells were no longer available (Supplementary Fig. 5d, absence of GFP signal) to supply labelled (blue) progeny to replenish the epithelium.

To study the long-term effects of CBC ablation, we isolated crypts from *Lgr5*^{DTR/+} mice to perform *in vitro* crypt organoid cultures¹⁴. Crypts depleted of *Lgr5*-expressing CBCs by treatment for 10 days with DT, as indicated by absence of GFP expression, gave rise to organoids with similar efficiency as wild-type controls (Supplementary Fig. 6a, b). These could be passaged *in vitro* in DT for up to 2 months without losing their ability to expand and proliferate. No *Lgr5*-expressing (GFP-positive) cells were detected in organoid epithelium as long as the organoids were maintained in medium containing DT (Supplementary Fig. 6d). However, when DT was removed from the culture medium, *Lgr5*-expressing cells reappeared at the bottom of crypt-like structures within 3 days (Supplementary Fig. 6c, GFP-positive cells).

Because we found that *Lgr5*-expressing CBCs were dispensable for crypt maintenance, we next asked whether *Bmi1*-expressing stem cells were mobilized to compensate for the loss of *Lgr5*-expressing stem cells. Mouse BMI1 regulates self-renewal of haematopoietic and neuronal stem cells¹⁵. We used a GFP knock-in allele (*Bmi1*^{GFP/+}) to monitor *Bmi1* gene expression¹⁶. *Bmi1*-expressing GFP-positive cells were most commonly observed at positions 3 to 6 above the crypt base (Fig. 3a), consistent with the *Bmi1* mRNA expression pattern in the small intestine². Upon depletion of *Lgr5*-expressing CBCs in *Lgr5*^{DTR/+}; *Bmi1*^{GFP/+} mice after 9 days of DT treatment, the number of GFP-positive cells per crypt increased three fold (Fig. 3a–d and Supplementary Fig. 7a), and the proportion of crypts containing either single or multiple GFP-expressing cells increased by 40% compared to control animals (Supplementary Fig. 7b). Of note, 55% of the total number of GFP-positive crypts in *Lgr5*^{DTR/+}; *Bmi1*^{GFP/+} mice now contained multiple GFP-positive cells (Fig. 3d and Supplementary Fig. 7b), compared with only 22% in *Bmi1*^{GFP/+} control animals.

To trace the fate of cells descended from *Bmi1*-expressing cells after elimination of *Lgr5*-expressing CBCs, we generated a *Bmi1*CreER bacterial artificial chromosome (BAC) transgenic allele (Supplementary Fig. 8). Labelling kinetics using the *Bmi1*-CreER transgenic line crossed with the R26R reporter were identical to previously reported results using the *Bmi1*^{CreER} knock-in allele² (Fig. 3f). *Bmi1*-CreER;R26R;*Lgr5*^{DTR/+} animals were treated with alternating doses of DT and TAM per day for 7 days (Fig. 3e). Because *Bmi1*-expressing cells are most abundant in the first 5 cm of the duodenum, we focused our analysis on this region. Consistent with the increased number of *Bmi1*-expressing cells (Supplementary Fig. 7a), the proportion of LacZ-positive crypts (either partially or fully labelled) also increased 34% upon loss of *Lgr5*-expressing CBCs (Supplementary Fig. 7c). The most marked difference was in the number of fully labelled crypts. Only 2.3% of crypts were fully labelled in *Bmi1*CreER;R26R control mice during a 6-day lineage tracing period, which was comparable with previous studies using a *Bmi1*^{CreER} knock-in allele². Upon loss of *Lgr5*-expressing CBCs, the number of fully labelled crypts increased approximately 15-fold (Fig. 3h, i and Supplementary Fig. 7c). These results indicate that in the absence of *Lgr5*-expressing cells, *Bmi1*-expressing cells are capable of directly giving rise to all intestinal cell types without going through *Lgr5*-positive intermediate cells. However, *Bmi1*-expressing stem cells did not give rise to an increased number of labelled crypts in more distal regions of small intestine and colon upon loss of *Lgr5*-expressing CBCs (Fig. 3f, g), indicating that alternative stem cell pools must compensate for the loss of *Lgr5*-expressing stem cells in distal regions of the gut.

Lastly, we tested whether *Bmi1*-expressing cells give rise to *Lgr5*-expressing cells under normal conditions. Because *Bmi1*- and *Lgr5*-expressing cells represent distinct although overlapping cell populations, we carried out a series of short-term pulse-chase experiments

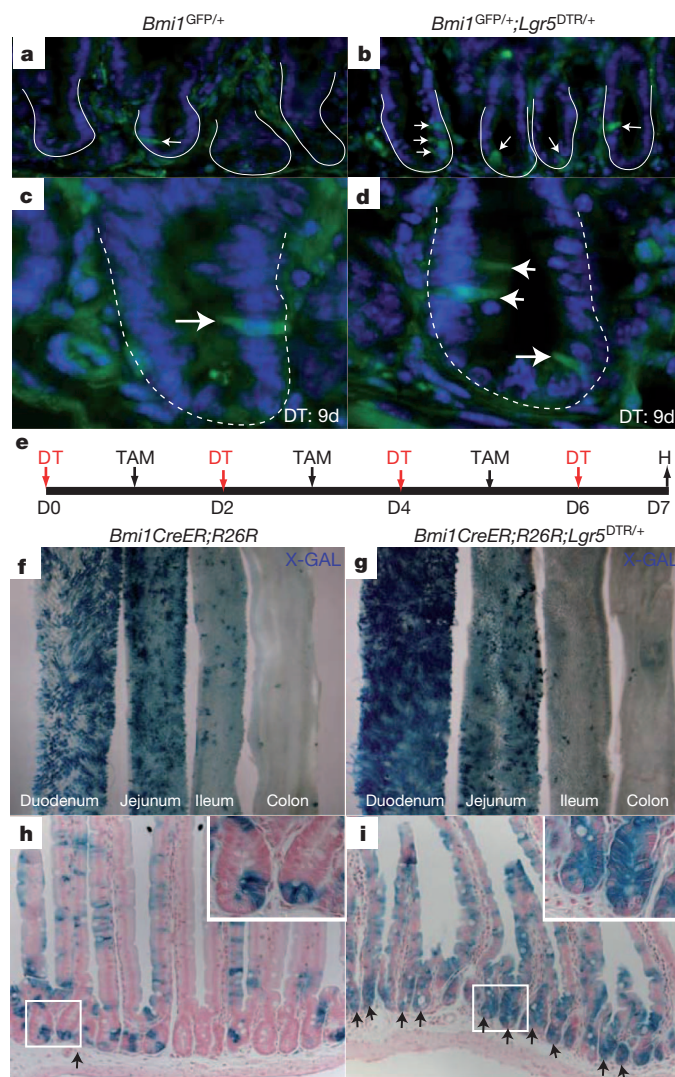


Figure 3 | *Bmi1*-expressing stem cells are mobilized to compensate for the loss of *Lgr5*-expressing CBCs. **a**, Rare *Bmi1*-expressing cells (arrows) are detected at positions 3 to 6 of the crypt base in the duodenum of *Bmi1*^{GFP/+} reporter mice. **b**, Increased *Bmi1*-expressing cells appear at the crypt base upon ablation of *Lgr5*-expressing CBCs. **c**, Higher magnification showing a *Bmi1*-expressing cell at position 4 of crypt base in *Bmi1*^{GFP/+} reporter mice. **d**, Close-up view of a crypt with multiple *Bmi1*-expressing cells after ablation of *Lgr5*-expressing cells. Arrows in **a–d** indicate (*Bmi1*)-expressing cells. **e**, Dosing regimen for lineage tracing of *Bmi1*-expressing cell progeny after ablation of *Lgr5*-expressing CBCs. **H**, harvest. **f, g**, Whole-mount X-GAL staining of the gastrointestinal tract. In both control mice and after ablation of *Lgr5*-expressing CBCs, *Bmi1*-expressing cells produce more progeny in the proximal than in the distal intestine. **h, i**, Close-up view of X-GAL-positive crypts in duodenum. Most of the labelled crypts have less than five X-GAL-positive cells in *Bmi1*-CreER;R26R control animals. Ablation of *Lgr5*-expressing CBCs stimulates production of progeny by *Bmi1*-expressing cells. 36% of the crypts in the first 5 cm of duodenum now become fully labelled (marked by arrows). Original magnification for panels: **a–d** at 40×; **f, g** at 1.2×; **h, i** at 20×.

using *Bmi1*-CreER;R26R;*Lgr5*^{DTR/+} mice. Twenty-four hours after TAM administration, most of the β-galactosidase (β-gal)-positive cells appeared as individuals, reflecting the normal pattern of *Bmi1* expression (Fig. 4a) in the initially labelled cells. *Bmi1*-expressing cells (β-gal positive) overlapped with *Lgr5*-expressing cells (GFP positive) between positions 1 to 6 at the crypt base; the double-positive cells peaked at positions 3 and 4 (Fig. 4a–c). This observation is consistent with a previous report stating that *Bmi1* mRNA expression (via quantitative polymerase chain reaction (qPCR) analysis) was readily detectable in *Lgr5*-positive cells¹¹. Later, between 48–72 h, clonal expansion

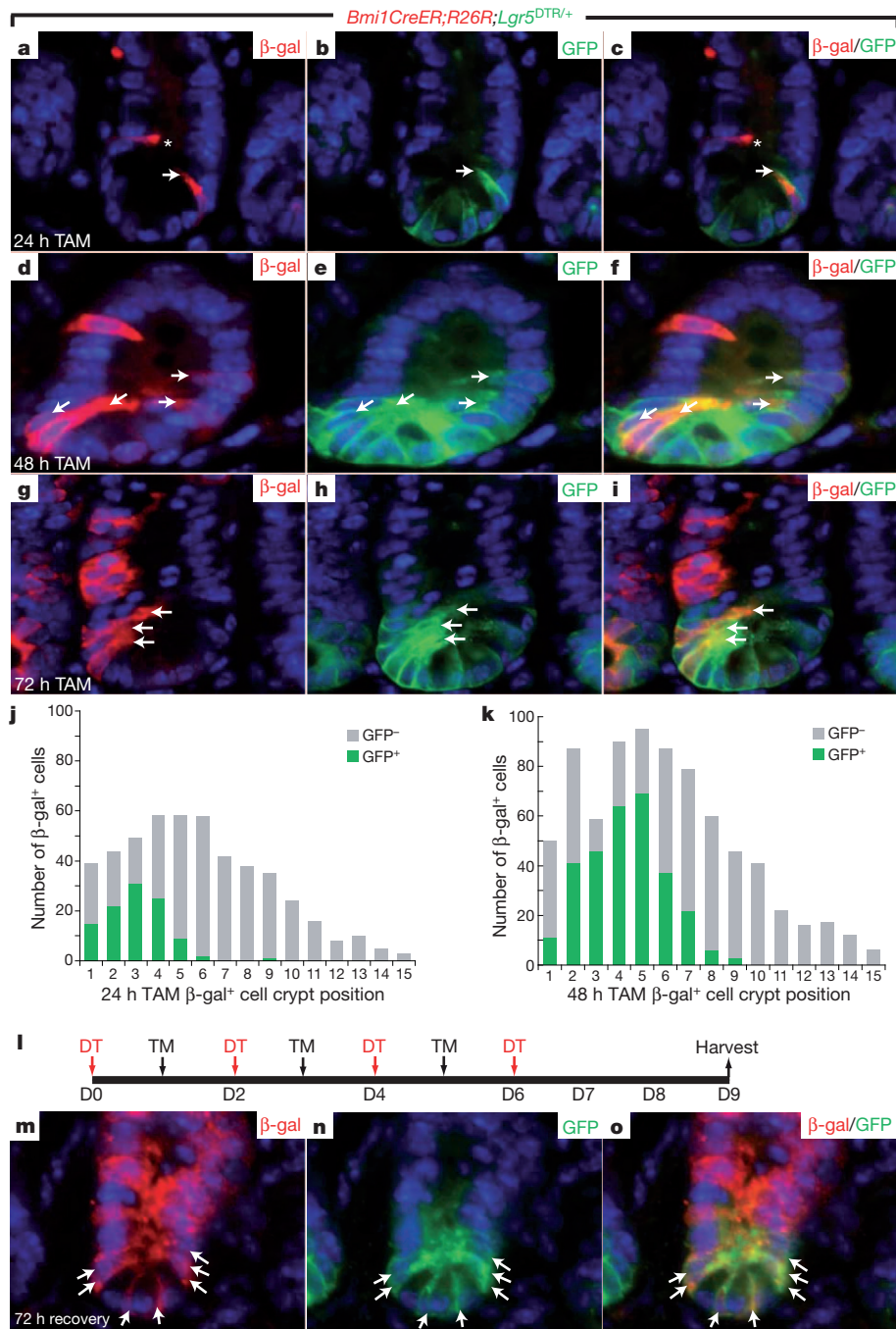


Figure 4 | *Bmi1*-expressing cells give rise to *Lgr5*-expressing CBCs under normal and injury conditions. **a–c**, *Bmi1*-*CreER*; *R26R*; *Lgr5*^{DTR/+} mice were dosed with 5 mg TAM and harvested 24 h later. β -Gal-positive cells (red) derived from *Bmi1*-expressing cells overlap with *Lgr5*-expressing CBCs (GFP-positive, green) at position 4 (arrow). A non-overlapping β -gal-positive cell was detected at position 7 in the same crypt (asterisk). **d–f**, More β -gal-positive cells (red) show overlapping expression (marked by arrow) with *Lgr5*-expressing CBCs (GFP⁺, green) at 48 h. **g–i**, At 72 h, clonal expansion from *Bmi1*-positive stem cells is now evident by a streak of β -gal-positive cells migrating upward (red). β -Gal-positive clones at lower crypt positions overlap with *Lgr5*-

expressing CBCs (arrow). **j, k**, Distribution of the *Bmi1*-positive stem cell progeny (β -gal⁺ cells) within the crypt at 24 and 48 h after TAM induction. **j**, *Bmi1*-expressing cells appear as singles throughout the crypt base between positions 1 to 15. **k**, More cells are derived from *Bmi1*-expressing stem cells at 48 h. A significant portion of β -gal⁺ cells also express *Lgr5* (GFP⁺, green column), at positions 1 to 6. Overlapping cells (green) peak around positions 3, 4 or 5. **l**, Dosing regimen used to study the recovery of *Lgr5*-expressing CBCs from *Bmi1*-positive cells. **m–o**, *Bmi1*-positive cells give rise to a fully labelled crypt (red), including newly formed *Lgr5*-expressing CBCs (GFP⁺, arrows). The original magnification for all panels is 40 \times .

from *Bmi1*-expressing cells was evident, as β -gal/GFP double-positive cells now appeared as doublets or triplets (Fig. 4d–i). We scored a total of 500 crypts at each time point and found that although a few cells were β -gal/GFP double positive (that is, expressing both *Bmi1* and *Lgr5*) at 24 h after TAM induction, this number doubled at 48 h (Fig. 4j, k). Similarly, lineage tracing from *Bmi1*-expressing cells carried out in mice treated for 6 days with DT and allowed to recover

for 72 h (Fig. 4l) demonstrated that newly formed *Lgr5*-positive cells at the bottom of the crypts arose from *Bmi1*-expressing cells (Fig. 4m–o). Together, these data show that *Bmi1*-expressing cells can give rise to *Lgr5*-expressing cells both under normal physiological conditions and after insults that deplete CBCs. Similar to our observation, *mTert*-expressing stem cells could also give rise to *Lgr5*-positive cells over a 5-day lineage tracing period⁹.

Our data support the existence of two stem cell pools in the epithelium of the small intestine: an actively proliferating stem cell compartment responsible for the daily maintenance of the intestine epithelium that is characterized by the expression of *Lgr5*, *Ascl2* and *Olfm4* (refs 1, 11, 17) and a distinct pool of stem cells expressing *Bmi1*. Our results lend support to the two-stem-cell pool model that is based on computational approaches¹⁸, and provide experimental evidence for recent models predicting that the intestine could fully recover after complete elimination of cellular subpopulations deemed to be functional stem cells¹⁹. Our data do not support the recent proposal that *Bmi1*-expressing cells are exclusively a subset of *Lgr5*-expressing cells¹¹; rather they indicate that under normal circumstances, *Bmi1*-positive stem cells are upstream of rapidly cycling, *Lgr5*-expressing stem cells and replenish the pool of active stem cells, either to avoid exhaustion of actively cycling stem cells or to prevent the accumulation of damaged cells that may lead to the development of tumours. Importantly, we also demonstrate that when the *Lgr5*-expressing cell compartment is eliminated by DT treatment, *Bmi1*-expressing cells can increase in number, presumably as a compensatory mechanism. Under these conditions, *Bmi1*-expressing cells contribute directly to the generation of all cell types of the intestinal epithelium to produce a functional organ until the rapidly cycling stem cell compartment is able to recover. Although it has been proposed that *Bmi1*-expressing stem cells are quiescent², this remains to be conclusively demonstrated.

Distinct stem cell pools with differing cycling dynamics have previously been observed in the hair follicle and in blood, organs that, like the intestine, undergo regular bouts of proliferation and regeneration^{20–22}. The factors that regulate the interplay between discrete populations of stem cells, and the precise hierarchical relationships among such populations, remain to be characterized. Although we have found that loss of *Lgr5*-positive cells is sustainable under short-term conditions *in vivo*, it remains to be determined whether such a scenario can persist for longer periods of time. Interestingly, depletion of Paneth cells, which are thought to be important for the maintenance of CBCs⁷, can be tolerated by mice for over 6 months without significant structural defects of the epithelium^{23,24}, supporting the idea that the intestine can function normally in the absence of CBCs. It will be important to determine how different stem cell populations sense the activity of other populations, whether rapidly cycling cells can repopulate more quiescent stem cell populations, and whether additional subpopulations of stem cells exist.

METHODS SUMMARY

Mice. *Lgr5*^{DTR/+}, *Lgr5*^{CreER/+} and *Bmi1*-*CreER* alleles were generated as described in Methods. *Bmi1*^{GFP/+} mice were provided by I. Weissman¹⁶. All studies and procedures involving animal subjects were approved by the Institutional Animal Care and Use Committees of Genentech and the University of California, San Francisco, and were conducted strictly in accordance with the approved animal handling protocol.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 April; accepted 1 August 2011.

Published online 18 September 2011.

1. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
2. Sangiorgi, E. & Capecchi, M. R. *Bmi1* is expressed *in vivo* in intestinal stem cells. *Nature Genet.* **40**, 915–920 (2008).

3. Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542–545 (2010).
4. Fuchs, E. The tortoise and the hair: slow-cycling cells in the stem cell race. *Cell* **137**, 811–819 (2009).
5. Zhu, L. *et al.* Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* **457**, 603–607 (2009).
6. Furuyama, K. *et al.* Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nature Genet.* **43**, 34–41 (2011).
7. Sato, T. *et al.* Paneth cells constitute the niche for *Lgr5* stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
8. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. *Am. J. Anat.* **141**, 537–561 (1974).
9. Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc. Natl Acad. Sci. USA* **108**, 179–184 (2011).
10. Muncan, V. *et al.* Rapid loss of intestinal crypts upon conditional deletion of the Wnt/Tcf-4 target gene *c-Myc*. *Mol. Cell. Biol.* **26**, 8418–8426 (2006).
11. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
12. Garcia, M. I. *et al.* LGR5 deficiency deregulates Wnt signaling and leads to precocious Paneth cell differentiation in the fetal intestine. *Dev. Biol.* **331**, 58–67 (2009).
13. Crosnier, C., Stamatakis, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature Rev. Genet.* **7**, 349–359 (2006).
14. Sato, T. *et al.* Single *Lgr5* stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
15. Park, I. K., Morrison, S. J. & Clarke, M. F. *Bmi1*, stem cells, and senescence regulation. *J. Clin. Invest.* **113**, 175–179 (2004).
16. Hosen, N. *et al.* *Bmi-1*-green fluorescent protein-knock-in mice reveal the dynamic regulation of *bmi-1* expression in normal and leukemic hematopoietic cells. *Stem Cells* **25**, 1635–1644 (2007).
17. van der Flier, L. G., Haeghebarth, A., Stange, D. E., van de Wetering, M. & Clevers, H. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology* **137**, 15–17 (2009).
18. Lobachevsky, P. N. & Radford, I. R. Intestinal crypt properties fit a model that incorporates replicative ageing and deep and proximate stem cells. *Cell Prolif.* **39**, 379–402 (2006).
19. Buske, P. *et al.* A comprehensive model of the spatio-temporal stem cell and tissue organisation in the intestinal crypt. *PLoS Comput. Biol.* **7**, e1001045 (2011).
20. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
21. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11**, 1351–1354 (2005).
22. Hsu, Y. C., Pasolli, H. A. & Fuchs, E. Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell* **144**, 92–105 (2011).
23. Bastide, P. *et al.* Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *J. Cell Biol.* **178**, 635–648 (2007).
24. Garabedian, E. M., Roberts, L. J., McNevin, M. S. & Gordon, J. I. Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. *J. Biol. Chem.* **272**, 23729–23740 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully acknowledge efforts by all the members of the Genentech mouse facility, in particular R. Ybarra and G. Morrow. We are grateful to N. Strauli, D.-K. Tran and A. Rathnayake for assistance with mouse breeding. We thank M. Roose-Girma, X. Rairdan and the members of the embryonic stem cell and microinjection groups for embryonic stem cell work and transgenic line generation and members of the F.J.d.S. laboratory for discussions and ideas. This work was funded in part by the National Institutes of Health through the NIH Director's New Innovator Award Program, 1-DP2-OD007191 and by R01-DE021420, both to O.D.K.

Author Contributions H.T., B.B., S.W., K.G.L. and L.R. designed, performed experiments and collected data. H.T., B.B., O.D.K. and F.J.d.S. designed experiments, analysed the data and wrote the manuscript. O.D.K. and F.J.d.S. are joint senior authors. All authors discussed results and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.J.d.S. (desauvage.fred@gene.com) or O.D.K. (ophir.klein@ucsf.edu).

METHODS

***Lgr5* and *Bmi1* vector construction.** The constructs for targeting the C57BL/6 *Lgr5* locus and the *Bmi1* BAC transgene were made using a combination of recombineering, DNA synthesis and standard molecular cloning techniques^{25,26}.

For *Lgr5*, a 7,213 bp fragment (assembly NCBI37/mm9, chr10:115,020,315–115,027,527) from a C57BL/6 BAC (RP23 library) was first retrieved into plasmid pBlight-TK²⁵. To generate the DTR–EGFP KI vector for *Lgr5*, a DTR–EGFP–pA–loxP–Neo–loxP cassette was synthesized (Blue Heron/Origene, DTR–EGFP sequence was based on that described previously²⁷, and inserted right after the ATG of *Lgr5* (chr10:115,024,547, reverse strand), deleting the remainder of exon 1 and splice donor of intron 1 (a 212 bp deletion). To generate the CreERT2 KI vector, a dsRed2–IRES–CreERT2–pA–Frt–neo–Frt cassette was synthesized (Blue Heron/Origene) and inserted at the same position as the DTR–EGFP cassette. The final vectors were confirmed by DNA sequencing.

The *Lgr5* KI vectors were linearized with NotI and C57BL/6 C2 embryonic stem cells were targeted using standard methods (G418-positive and gancyclovir-negative selection). Positive clones were identified using PCR and taqman analysis, and confirmed by sequencing of the modified locus. Correctly targeted embryonic stem cells were transfected with a Cre or Flpe plasmid, respectively, to remove the Neo cassette. The modified embryonic stem cells were then injected into blastocysts using standard techniques, and germline transmission was obtained after crossing the resulting chimaeras with C57BL/6N females.

For *Bmi1*, a 210 kb C57BL/6 BAC (RP23-181D14, assembly NCBI37/mm9, chr2:18,464,619–18,674,471) was obtained and characterized by DNA fingerprinting. The BAC contains the *Bmi1* locus and considerable 5' and 3' flanking sequence. An IRES–CreERT2–pA–Frt–Neo–Frt cassette was synthesized (Blue Heron/Origene) and inserted, using recombineering, 85 bp 3' of the *Bmi1* stop codon (after position chr2:18,606,193). Neo was then removed by transforming the modified BAC into arabinose-induced, SW105 cells^{28,29} expressing the yeast protein Flp. C57BL/6 transgenic mice carrying the modified *Bmi1* BAC were obtained using standard pronuclear microinjection methods³⁰ and characterized.

We analysed the *Lgr5*^{DTR/+} mice at 24 h after DT administration (50 µg kg⁻¹, intraperitoneal injection, *n* = 3), at 10 days of DT treatment (50 µg kg⁻¹ every other day for 10 days, *n* = 5), 48 h recovery (*n* = 3) and 96 h recovery time points after four doses of DT (*n* = 2). The DT treatment could not be extended beyond 10 days due to severe liver toxicity apparently mediated by a subset of *Lgr5*–DTR–EGFP-expressing hepatocytes. We analysed *Bmi1*CreER;R26R;*Lgr5*^{DTR/+} at 24 h (*n* = 2), 48 h (*n* = 2) and 72 h (*n* = 2) after TAM injection. Two-hundred and fifty β-gal-positive crypts were scored per mouse.

Renal capsule explants. 3–5-mm small intestine pieces from E15 *Lgr5*^{DTR/CreER} embryos (*n* = 3) were grafted under the renal capsule of 6–8-week-old athymic nu/nu mice and allowed to develop for 3 weeks. We treated *Lgr5*^{DTR/CreER};R26R renal grafts with 10 days TAM (*n* = 5), 10 days DT/TAM (*n* = 5), 6 days DT followed by 6 days TAM (*n* = 5) and 6 days TAM followed by 6 days DT

(*n* = 5). Some GFP expression was seen outside of the CBC region due to perdurance of GFP protein as well as upregulation of the *Lgr5* locus when *Lgr5* is deleted¹². **DT cell ablation.** Mice (between 6 and 12 weeks old) were given DT at 50 µg kg⁻¹ every 48 h through intraperitoneal injections.

TAM labelling experiments. Mice (between 6 and 12 weeks old) were given 5 mg TAM in corn oil through intraperitoneal injection.

Transmission electron microscopy. The tissues were fixed in 1/2 Karnovsky's fixative (2% paraformaldehyde (PFA), 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.2), washed in the same buffer, and post-fixed in 1% osmium tetroxide. The samples were then dehydrated through a series of ethanol, followed by propylene oxide and embedded in Eponate 12 (Ted Pella). Thin sections were stained with uranyl acetate and lead citrate and examined using a Philips CM12 or JEOL JEM-1400 TEM.

Histology, immunohistochemistry and immunofluorescence. Animals were perfused with 2% PFA. Small intestine and colon were flushed with 2% PFA and fixed in 4% PFA overnight. Half of the materials were cryo-protected, embedded in OCT, and sectioned at 6 µm for immunofluorescence. The other half of the materials were paraffin embedded, sectioned at 3 µm for histology and immunohistochemistry. Antibodies: Ki67 (Neomarker), cleaved caspase 3 (Cell Signaling), GFP (Novus), chromogranin A (Neomarkers), β-gal (Cappel).

In situ hybridization and X-GAL staining. Full-length *Lgr5* cDNA was cloned into the pGEM vector to make anti-sense DIG-probe. Protocols for *in vitro* transcription and *in situ* hybridization were as described previously³¹. Whole-mount X-GAL staining was performed as described³.

Crypt organoid culture. Crypt isolation and culture were performed as described¹⁴.

25. Warming, S., Rachel, R. A., Jenkins, N. A. & Copeland, N. G. *Zfp423* is required for normal cerebellar development. *Mol. Cell. Biol.* **26**, 6913–6922 (2006).
26. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
27. Kissenpfennig, A. *et al.* Dynamics and function of Langerhans cells *in vivo*: dermal dendritic cells colonize lymph node areas distinct from slower migrating Langerhans cells. *Immunity* **22**, 643–654 (2005).
28. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).
29. Lee, E. C. *et al.* A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56–65 (2001).
30. Van Keuren, M. L., Gavrilina, G. B., Filipiak, W. E., Zeidler, M. G. & Saunders, T. L. Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic Res.* **18**, 769–785 (2009).
31. Gregorieff, A. & Clevers, H. *In situ* hybridization to identify gut stem cells. *Curr. Protoc. Stem Cell Biol.* Ch. 2, Unit 2F.1 (2010).

Mechanical strain in actin networks regulates FilGAP and integrin binding to filamin A

A. J. Ehrlicher^{1,2}, F. Nakamura¹, J. H. Hartwig¹, D. A. Weitz² & T. P. Stossel¹

Mechanical stresses elicit cellular reactions mediated by chemical signals. Defective responses to forces underlie human medical disorders^{1–4} such as cardiac failure⁵ and pulmonary injury⁶. The actin cytoskeleton's connectivity enables it to transmit forces rapidly over large distances⁷, implicating it in these physiological and pathological responses. Despite detailed knowledge of the cytoskeletal structure, the specific molecular switches that convert mechanical stimuli into chemical signals have remained elusive. Here we identify the actin-binding protein filamin A (FLNA)^{8,9} as a central mechanotransduction element of the cytoskeleton. We reconstituted a minimal system consisting of actin filaments, FLNA and two FLNA-binding partners: the cytoplasmic tail of β -integrin, and FilGAP. Integrins form an essential mechanical linkage between extracellular and intracellular environments, with β -integrin tails connecting to the actin cytoskeleton by binding directly to filamin⁴. FilGAP is an FLNA-binding GTPase-activating protein specific for RAC, which *in vivo* regulates cell spreading and bleb formation¹⁰. Using fluorescence loss after photoconversion, a novel, high-speed alternative to fluorescence recovery after photobleaching¹¹, we demonstrate that both externally imposed bulk shear and myosin-II-driven forces differentially regulate the binding of these partners to FLNA. Consistent with structural predictions, strain increases β -integrin binding to FLNA, whereas it causes FilGAP to dissociate from FLNA, providing a direct and specific molecular basis for cellular mechanotransduction. These results identify a molecular mechanotransduction element within the actin cytoskeleton, revealing that mechanical strain of key proteins regulates the binding of signalling molecules.

The composite cytoskeleton network *in vivo* provides dynamic cellular structure and actively generates movement. A physiological reconstituted *in vitro* network of actin and FLNA creates an elastic gel mechanically dominated by the rod-like actin filaments and cross-linked by flexible FLNA molecules. Applying strain to this network readily deforms FLNA crosslinks (Fig. 1a, b), and the specific structure and actin binding of FLNA suggest how these deformations might affect FLNA's interaction of with some of its 90 or so other currently identified binding partners⁹.

FLNA is an extended homodimer composed of two identical subunits, each having an amino-terminal actin-binding domain followed by 24 immunoglobulin repeats¹² (Fig. 1c, d). The actin-binding domains and repeats 1–15 are designated 'rod 1' and form a linear structure that binds actin filaments. Repeats 16–23, comprising 'rod 2', however, form compact globular clusters that do not interact with actin filaments and contain most of FLNA's binding-partner sites. Strain-dependent reversible straightening of these domains contributes to FLNA-actin network flexibility and may regulate local binding-partner affinity (Supplementary Fig. 1). Here we examine the effects of mechanical strain on FLNA's interactions with two key rod-2 binding partners; cytoplasmic β -tail integrin, which nucleates an extensively characterized signalling¹³ and adhesion¹⁴ complex, and FilGAP, which is a GTPase specific for RAC, a regulator of cellular activity such as actin

assembly¹⁰. Mechanical strain may regulate partner binding, and we propose that stretching FLNA crosslinks not only causes FilGAP to unbind, but also causes integrin to bind more strongly (Fig. 1c, d and Supplementary Fig. 1). Neighbouring immunoglobulin repeats cover integrin binding sites on FLNA repeats 19 and 21 (refs 15, 16), but computational simulations suggest that rod 2 of FLNA is highly flexible and that physiological forces are sufficient to expose these cryptic sites, allowing integrin to bind^{17,18} (Supplementary Fig 1a, b). FilGAP binding occurs on each repeat 23, suggesting that FilGAP is able to bind repeat 23 on both subunits simultaneously when unstressed, providing sufficient avidity to promote FilGAP association with FLNA (Fig. 1c and Supplementary Fig. 1c). Mechanical stretching of FLNA spatially separates repeats 23, preventing FilGAP from binding simultaneously to both¹⁹, thus causing it to dissociate (Fig. 1d and Supplementary Fig. 1d).

To test these hypotheses and measure the effect of mechanical stress on binding-partner interactions with FLNA, we reconstituted networks of polymerized actin (F-actin) and FLNA containing the binding-partner FilGAP or β_7 -integrin. To quantify the strain-dependent

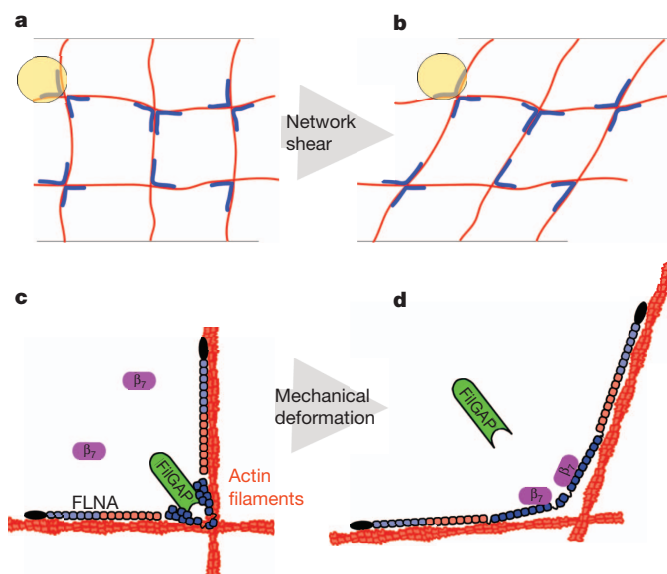


Figure 1 | Differential mechanotransduction in FLNA occurs through spatial separation of binding sites and opening cryptic sites. **a**, A gel of actin (red) crosslinked with filamin (blue) forms an orthogonal network. **b**, When this network is strained, crosslinks are deformed. **c**, The actin-binding domain of FLNA is shown in black, and is followed by repeats 1–7 (light blue) and 8–15 (red/orange), which form the linear rod-1 region. Repeats 16–23 (dark blue) form the compact rod-2 region. FilGAP (green) binds repeats 23 and the cytoplasmic domain of β_7 -integrin (purple) is unbound. **d**, When FLNA is mechanically deformed, the cryptic integrin site on repeat 21 is exposed allowing β_7 -integrin to bind, whereas repeats 23 are spatially separated, preventing FilGAP from binding them both.

¹Translational Medicine Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ²School of Engineering and Applied Sciences, Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

kinetics of these partners of FLNA, we developed fluorescence loss after photoconversion (FLAC), which takes advantage of the rapid photoactivation or conversion of photoactivatable fluorescent proteins (PAFPs) as a high-speed analogue to fluorescence recovery after photobleaching¹¹. In FLAC, a sample with an initially non-fluorescent binding partner is locally illuminated with a 50-ms pulse of 405-nm light, rapidly and permanently activating PAFP-conjugated partner fluorescence (Supplementary Figs 4 and 5). Photoactivation fluorescently marks the sample faster than conventional photobleaching, and without the requirement of high excitation flux. After activation, unbound PAFP rapidly diffuses away, decreasing the fluorescence signal, whereas bound PAFP dissociates more slowly. The time-dependent decay of PAFP intensity reveals the kinetics of the FLNA binding partner, as slower decay indicates slower unbinding, and thus provides a direct, high-speed assay of dissociation.

We tested the utility of these PAFP constructs in assaying binding kinetics by reconstituting F-actin and the PAFP-labelled binding partner with different forms of FLNA that have higher or lower affinities for β_7 -integrin or FilGAP. Consistent with immunoprecipitation data (Supplementary Fig. 3b, c), the fluorescence decay of β_7 -integrin labelled with photoactivatable green fluorescent protein (PA-GFP β_7 -integrin) was faster in wild-type FLNA networks than in the del41 variant (Supplementary Movie 1), demonstrating relatively stronger binding in the del41 mutant than in wild type. The fluorescence decay of PA-GFP FilGAP was slower in wild-type FLNA networks than in the M2474E mutant (Supplementary Movie 2), also in agreement with immunoprecipitation data (Supplementary Fig. 3a).

We then applied the FLAC methodology to measure the mechanosensitive aspect of interactions between PAFP-labelled binding partners and FLNA. We sheared networks of F-actin and FLNA containing PAFP-tagged FilGAP or β_7 -integrin in a precise and highly controlled fashion using a microscope stage comprising a stationary coverslip for the bottom of the sample and a piezo-controlled linear actuator for the top. When the FLNA/F-actin network was not strained, β_7 -integrin had a characteristic exponential decay time of 1.3 ± 0.1 s. The application of a shear strain, $\gamma = 0.28$, increased this time to 3.5 ± 0.3 s (Fig. 2a). The change in fluorescence decay rate under strain describes how the geometric state of FLNA affects dissociation of β_7 -integrin; thus, mechanically stretching FLNA molecules enhanced the β_7 -integrin binding. By contrast, FilGAP behaved qualitatively oppositely: unstrained networks had a characteristic fluorescence decay time of 3.6 ± 0.7 s, which decreased to 0.6 ± 0.1 s when a shear strain of $\gamma = 0.28$ was applied (Fig. 2b). FLNA does not permanently crosslink actin, and by unbinding and rebinding on the timescale of ~ 6 min (Supplementary Fig. 6), it dynamically allows the network to relax to an unstressed state. After 10 min under strain, the network had sufficient time to dissipate

internal stress through FLNA remodelling, and the fluorescence decay time increased to 6.1 ± 0.7 s, demonstrating the reversibility of strain-modulated FilGAP binding to FLNA (Fig. 2b).

The application of unidirectional shear reveals the effects of strain on partner binding to FLNA; however, cells commonly generate internal stresses using molecular motors such as myosin. To examine the effects of cytoskeleton-induced stress, and as a physiological technique complementary to external shear, we included myosin II in the networks to generate contractile stress²⁰ (Supplementary Fig. 9 and Supplementary Movie 3). We allowed the composite network to assemble and come to an unstressed equilibrium state over ~ 6 h, after which time the incorporated myosin II had ceased contracting, by enzymatically exhausting the pool of added ATP, and dynamic FLNA remodelling had dissipated internal stresses. For unstressed FLNA, we measured β_7 -integrin and FilGAP fluorescence decay times of 1.6 ± 0.1 and 1.5 ± 0.1 s, respectively (Fig. 3a, c). Including photo-labile ‘caged’ ATP in the sample allowed us to release fresh ATP and restart myosin motor activity^{21,22}, which contracts the actin network and strains FLNA crosslinks. In myosin-stressed FLNA, the integrin unbinding time increased to 2.5 ± 0.2 s but the FilGAP unbinding time decreased to 0.9 ± 0.1 s (Fig. 3b, d). The application of either external shear or myosin contraction resulted in increased integrin binding and decreased FilGAP binding, demonstrating the robust, opposite behaviours of these FLNA binding partners.

The FLNA-crosslinked actin cytoskeleton is a large, percolated network that, owing to its filamentous actin structure, can readily transmit large mechanical deformations over long intracellular distances, yet FLNA is mechanosensitive to nanometre-scale molecular deformations. This range of length scales contrasts with that of focal adhesion mechanosensitivity, which detects local mechanics and is limited to small spatial and strain scales owing to their size and connectivity^{23,24}.

In conclusion, we have developed *in vitro* systems to study quantitatively protein–protein interactions under mechanical force. Using PAFPs with the FLAC technique provides the increases in time resolution necessary for measuring transient kinetics, without the harsh intensity or duration of bleaching exposure required for fluorescence recovery after photobleaching. The results presented here establish FLNA as a mechanotransductive substrate within the cytoskeleton and highlight the utility of *in vitro* systems, in combination with FLAC, to determine quantitative responses of specific proteins.

Mechanotransduction *in vitro* provides the biological specificity necessary for understanding how these complex regulatory signals may operate *in vivo*. Cellular mechanotransduction has been shown to induce rapid biochemical activity over long distances²⁵. Because mechanical stimuli induce relatively large local deformations that decrease in magnitude with distance from the site of application,

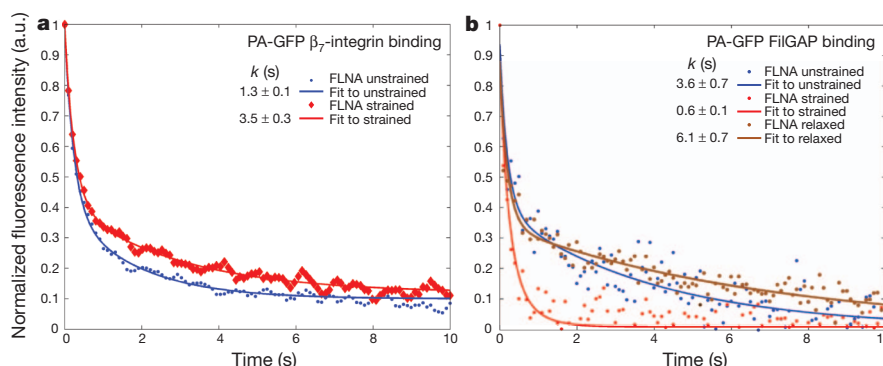


Figure 2 | External bulk shear of F-actin/FLNA networks alters FLNA's binding affinity for β_7 -integrin and FilGAP. **a**, Fluorescence intensity of PA-GFP β_7 -integrin as a function of time after photoactivation. When unstrained (blue), fluorescence of β_7 -integrin decays with a characteristic time constant of $k = 1.3$ s. Following the application of a shear strain of $\gamma = 0.28$, the time constant increases to 3.5 s, as the integrin dissociates more slowly from FLNA

($n = 18$). a.u., arbitrary units. **b**, Fluorescence intensity of PA-GFP FilGAP as a function of time after photoactivation. Unstrained (blue) FilGAP's fluorescence decay time is $k = 3.6$ s. A shear strain of $\gamma = 0.28$ (red) decreased k to 0.6 s. This behaviour is reversible, and after allowing the network to relax for 10 min, removing all strain, k increases to 6.1 s (brown) ($n = 10$).

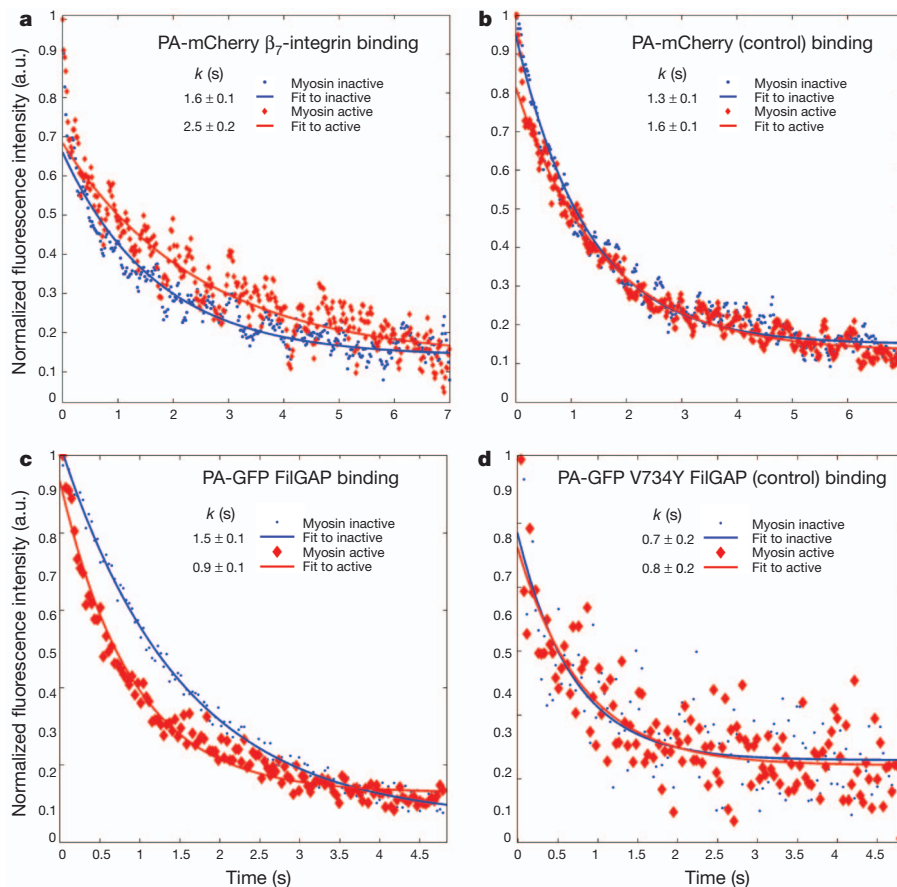


Figure 3 | Myosin II forces applied to F-actin/FLNA networks change FLNA's binding affinity to β_7 -integrin and FilGAP. **a**, When depleted of ATP, myosin is in a rigor state. The FLNA within the network is not stressed and PA-GFP β_7 -integrin fluorescence decays with a characteristic time constant of $k = 1.6$ s (blue). After caged ATP is released, myosin reactivates, straining FLNA crosslinks. The decay time constant increases to 2.5 s because the integrin dissociates more slowly from FLNA under stress. **b**, PA-mCherry

FLNA mechanotransduction *in vivo* probably provides a rapid, distance-sensitive biphasic response by binding or unbinding integrins or FilGAP, respectively, as a result of the transmitted strain. Physiologically, the localization and binding of these proteins determine their activity. Strain-induced binding of integrin to FLNA may compete with talin binding to integrin²⁶, thus providing a mechanosensitive switch for integrin activation and adhesion. FLNA's homodimer structure may induce clustering of integrin, thereby reinforcing adhesion and concentrating signalling molecules at a specific location. FilGAP, when unbound from FLNA, relocates to the plasma membrane, where it inactivates RAC¹⁰. Active RAC levels profoundly affect cell movement²⁷, and increased RAC activity in FLNA-deficient cells correlates with increased apoptosis²⁸. Moreover, our measurements are consistent with *in vivo* studies demonstrating that RAC activity and expression seem to be force-regulated by FilGAP–FLNA interactions, because inhibiting FLNA or FilGAP increases RAC activity but applying local forces to wild-type cells causes FilGAP to decrease RAC activity²⁸. Because FLNA does not change FilGAP's catalytic activity, mechanically induced redistribution alone might explain its regulation *in vivo*. Force-dependent conformational changes in structure required for mechanical regulation have been observed in many proteins, including FLNA *in vivo*^{29,30}. By identifying FLNA as a mechanosensitive element within the cytoskeleton, we have clarified how RAC and integrin activity may be regulated by a specific molecular mechanotransduction pathway. Identifying mechanotransduction elements may direct novel, targeted therapeutic approaches by correcting or modulating mechanosensitive binding.

alone as a control shows no significant difference between the unstrained and strained states. **c**, Fluorescence intensity of PA-GFP FilGAP as a function of time after photoactivation. **d**, In the ATP-depleted state, FilGAP's fluorescence decay time is $k = 1.5$ s, and after release of the caged ATP (red), k decreases to 0.9 s. In PA-GFP V734Y FilGAP, a non-FLNA binding mutant used as a control, the decay time in the unstrained state (0.7 s) is not significantly different from that in the strained state (0.8 s) ($n = 20$).

METHODS SUMMARY

PAFP fluorophore synthesis. PAFP fluorophore was genetically tagged to binding partners, creating PAFP-labelled β_7 -integrin and FilGAP. Solubility and correct binding of labelled partners was confirmed using western blots (Supplementary Fig. 3).

FLAC methodology. An external 405-nm laser was coupled into a Leica SP5 confocal microscope and used to illuminate a central, ~ 2 - μ m spot for 50 ms, converting the PAFP from its dark state to its fluorescent state (Supplementary Figs 4 and 5). The decay in the fluorescence intensity, $I(t)$, of the activated fluorophores was monitored and fitted with the exponential $I(t) = ae^{-t/k} + c$, where k is the time constant of characteristic dissociation. Given k values represent best fits plus 95% confidence intervals. In the case of single-step uniaxial strain presented in Fig. 2, data were fitted to $I(t) = ae^{-t/k} + 0.5e^{-t/(0.15s)} + c$ to provide a more accurate fit and compensate for the rapid diffusion of free fluorophore.

Sample cell composition. Shear cell samples consisted of 24 μ M actin, 0.12 μ M FLNA, $\times 1$ F-actin polymerization buffer (Methods), 2 μ M Alexa 546 phalloidin and either PA-GFP FilGAP or PA-GFP β_7 -integrin, and were sheared in a piezo-driven shear cell (Supplementary Information). Sheared FLAC measurements for strained networks were acquired approximately 5–10 s after shear. Myosin samples included 24 μ M actin, 0.12 μ M FLNA, 1 μ M myosin II, $\times 1$ activity buffer (Methods) and 2 μ M caged ATP, along with 2 μ M Alexa 546 phalloidin and PA-GFP FilGAP or 2 μ M Alexa 488 phalloidin and PA-mCherry β_7 -integrin. Samples were allowed to polymerize and consume available ATP over ~ 6 h. FLAC measurements were then performed on the ATP-free unstrained network. Subsequently, the caged ATP (Sigma) was released by a 4-s exposure to a diffuse 50-mW, 365-nm light-emitting diode (Prizmatix), and within 3 s the network could be seen to homogenize under myosin contraction (Supplementary Fig. 9 and Supplementary Movie 3). FLAC measurements were then repeated in this active myosin-stressed network to quantify the strain-dependent binding activity.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 April; accepted 5 August 2011.

Published online 18 September 2011.

1. Ingber, D. E. Mechanobiology and diseases of mechanotransduction. *Ann. Med.* **35**, 564–577 (2003).
2. Discher, D. E., Janmey, P. & Wang, Y. L. Tissue cells feel and respond to the stiffness of their substrate. *Science* **310**, 1139–1143 (2005).
3. Moore, S. W., Roca-Cusachs, P. & Sheetz, M. P. Stretchy proteins on stretchy substrates: the important elements of integrin-mediated rigidity sensing. *Dev. Cell* **19**, 194–206 (2010).
4. Hoffman, B. D. *et al.* Dynamic molecular processes mediate cellular mechanotransduction. *Nature* **475**, 316–323 (2011).
5. Krüger, M. & Linke, W. A. Titin-based mechanical signalling in normal and failing myocardium. *J. Mol. Cell. Cardiol.* **46**, 490–498 (2009).
6. Birukov, K. G. Small GTPases in mechanosensitive regulation of endothelial barrier. *Microvasc. Res.* **77**, 46–52 (2009).
7. Wang, N., Tytell, J. D. & Ingber, D. E. Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus. *Nature Rev. Mol. Cell Biol.* **10**, 75–82 (2009).
8. Stossel, T. P. *et al.* Filamins as integrators of cell mechanics and signalling. *Nature Rev. Mol. Cell Biol.* **2**, 138–145 (2001).
9. Nakamura, F., Stossel, T. P. & Hartwig, J. H. The filamins: organizers of cell structure and function. *Cell Adh. Migr.* **5**, 160–169 (2011).
10. Ohta, Y., Hartwig, J. H. & Stossel, T. P. FilGAP, a Rho- and ROCK-regulated GAP for Rac binds filamin A to control actin remodelling. *Nature Cell Biol.* **8**, 803–814 (2006).
11. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* **86**, 3473–3495 (2004).
12. Nakamura, F., Osborn, T. M., Hartemink, C. A., Hartwig, J. H. & Stossel, T. P. Structural basis of filamin A functions. *J. Cell Biol.* **179**, 1011–1025 (2007).
13. Wang, N., Butler, J. P. & Ingber, D. E. Mechanotransduction across the cell surface and through the cytoskeleton. *Science* **260**, 1124–1127 (1993).
14. Calderwood, D. A. *et al.* Increased filamin binding to β -integrin cytoplasmic domains inhibits cell migration. *Nature Cell Biol.* **3**, 1060–1068 (2001).
15. Lad, Y. *et al.* Structure of three tandem filamin domains reveals auto-inhibition of ligand binding. *EMBO J.* **26**, 3993–4004 (2007).
16. Heikkinen, O. K. *et al.* Atomic structures of two novel immunoglobulin-like domain pairs in the actin cross-linking protein filamin. *J. Biol. Chem.* **284**, 25450–25458 (2009).
17. Pentikäinen, U. & Ylanne, J. The regulation mechanism for the auto-inhibition of binding of human filamin A to integrin. *J. Mol. Biol.* **393**, 644–657 (2009).
18. Chen, H. S., Kolahi, K. S. & Mofrad, M. R. Phosphorylation facilitates the integrin binding of filamin under force. *Biophys. J.* **97**, 3095–3104 (2009).
19. Nakamura, F. *et al.* Molecular basis of filamin A-FilGAP interaction and its impairment in congenital disorders associated with filamin A mutations. *PLoS ONE* **4**, e4928 (2009).
20. Koenderink, G. H. *et al.* An active biopolymer network controlled by molecular motors. *Proc. Natl Acad. Sci. USA* **106**, 15192–15197 (2009).
21. Humphrey, D., Duggan, C., Saha, D., Smith, D. & Kas, J. Active fluidization of polymer networks through molecular motors. *Nature* **416**, 413–416 (2002).
22. Smith, D. M. *et al.* Molecular motor-induced instabilities and crosslinkers determine biopolymer organization. *Biophys. J.* **93**, 4445–4452 (2007).
23. Kanchanawong, P. *et al.* Nanoscale architecture of integrin-based cell adhesions. *Nature* **468**, 580–584 (2010).
24. Grashoff, C. *et al.* Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature* **466**, 263–266 (2010).
25. Na, S. *et al.* Rapid signal transduction in living cells is a unique feature of mechanotransduction. *Proc. Natl Acad. Sci. USA* **105**, 6626–6631 (2008).
26. Kiema, T. *et al.* The molecular basis of filamin binding to integrins and competition with talin. *Mol. Cell* **21**, 337–347 (2006).
27. Sanz-Moreno, V. *et al.* Rac activation and inactivation control plasticity of tumor cell movement. *Cell* **135**, 510–523 (2008).
28. Shifrin, Y., Arora, P. D., Ohta, Y., Calderwood, D. A. & McCulloch, C. A. The role of FilGAP-filamin A interactions in mechanoprotection. *Mol. Biol. Cell* **20**, 1269–1279 (2009).
29. Johnson, C. P., Tang, H. Y., Carag, C., Speicher, D. W. & Discher, D. E. Forced unfolding of proteins within cells. *Science* **317**, 663–666 (2007).
30. Krieger, C. C. *et al.* Cysteine shotgun-mass spectrometry (CS-MS) reveals dynamic sequence of protein structure changes within mutant and stressed cells. *Proc. Natl Acad. Sci. USA* **108**, 8269–8274 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors acknowledge the Harvard Materials Research and Engineering Center (DMR-0820484) for confocal imaging, and M. Ginsberg, J. Lippincott-Schwartz and V. Verkhusha for providing complementary DNA for PA-GFP and PA-mCherry constructs. We thank T. Collins for technical assistance, L. Jawerth and V. Zaburdaev for discussions, and J. Wilking and K. Guentner for help with the manuscript. This work was supported by grants NIH R01 HL19429 (T.P.S.) and NIH T32 HL07680 (A.J.E.) and by the Harvard University Science and Engineering Committee Seed Fund for Interdisciplinary Science (D.A.W., T.P.S., F.N.).

Author Contributions The project was initiated by T.P.S., F.N. and D.A.W. Experiments were designed by A.J.E., T.P.S., J.H.H. and F.N. Proteins and materials were synthesized and purified by F.N. and A.J.E. FLAC experiments were performed by A.J.E. and binding assays were performed by F.N. Data was analysed by A.J.E. All authors discussed data and aided in preparing the manuscript. A.J.E. and F.N. contributed equally to this project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.N. (fnakamura@rics.bwh.harvard.edu) or A.J.E. (aje@seas.harvard.edu).

METHODS

Protein design and synthesis. Actin was purified from rabbit skeletal muscle and gel-filtered (HiLoad 16/60 Superdex 200pg; GE Healthcare) in G buffer¹² (2 mM Tris-HCl, 0.2 mM ATP, 0.2 mM CaCl₂, 0.2 mM DTT and 0.005% NaN₃, pH 8.0). Aliquots of purified G-actin were frozen in liquid nitrogen and stored at -80°C . Before use, G-actin was thawed 12 h in advance and dialyzed against fresh G buffer. Myosin II from rabbit skeletal muscle was obtained from Cytoskeleton (Denver). Human full-length FLNA and FilGAP were expressed using a baculovirus expression system (Invitrogen) in sf9 insect cells and purified as previously described¹⁹. All the point or deletion mutants were generated using the QuickChange site-directed mutagenesis kit (Agilent Technologies). FilGAP and integrin constructs were expressed in sf9 cells as follows. The complementary DNA (cDNA) encoding PAFPs (PA-GFP and PA-mCherry, which were kind gifts from Jennifer Lippincott-Schwartz, NIH) were amplified by PCR using the forward primer GAAGATCTATGGTGAGCAAGGGCGAGG and the reverse primer CGGGATCCCTTGTACAGCTCGTCCATG, and introduced into BamHI sites of pFASTBAC-HTb vector¹² to construct pFASTBAC-HTb-PAFPs. The cDNA encoding the cytoplasmic domain (amino acids 769–789) of human β_7 -integrin was amplified by PCR using pET15-G4-integrin β_7 (a kind gift from Mark Ginsberg, UC San Diego) as a template with the forward primer GCGGATCCAACTGGAAGCAGGACAGTAATC and the reverse primer CGGAATTCAGCGAGGATTGATGGTGG, and inserted into BamHI/EcoRI sites of pFASTBAC-HTb-PAFPs. For FilGAP, the cDNA encoding PAFPs was introduced into pFASTBAC-HTa-FilGAP at the XbaI cleavage site¹⁹. The expressed proteins were purified by Ni-NTA affinity and gel filtration chromatography (Superose 12 and Superdex 200, GE Healthcare) as previously described¹⁹. Protein concentration was measured by absorption at 280 nm using parameters calculated with ProtParam (<http://au.expasy.org/tools/protparam.html>). Genetic fusion of PAFPs to the binding partners³¹ did not affect the binding-partner activity. Purified PAFP FilGAP proteins interact with full-length FLNA in the same dose-dependent manner as unlabelled FilGAP, and do not bind to the FLNA M2474E mutant, which lacks the FilGAP-binding site (Supplementary Fig. 2a). The PAFP-tagged cytoplasmic tail of β_7 -integrin was also found to retain its binding behaviour with FLNA, and predominantly interacts with the FLNA del41 variant, where the autoinhibitory ligand-binding site is constitutively exposed^{15,32} (Supplementary Fig. 2b, c).

Protein pull-down assay. The purified His-FilGAP and His-PAFP FilGAP constructs were incubated with increasing amounts of FLAG-FLNA and 20 μl of FLAG-specific mAb M2 agarose (50% (v/v) slurry, Sigma) in binding buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% (w/v) Triton X-100, 0.1 mM β -mercaptoethanol, 0.1 mM EGTA, pH 7.4; 400 μl) for 1 h at 25°C . The beads were sedimented and washed three times with binding buffer. Proteins bound to the beads were solubilized in SDS sample buffer and separated by 9.0% (w/v) SDS-PAGE followed by immunoblotting using rabbit polyclonal antibodies (pAbs) against FilGAP¹⁰. For integrin, the purified His-PAFP β_7 -integrin (amino acids 769–789) was incubated with increasing amounts of wild-type and mutant (del41; amino-acid residues 2126–2167 are deleted) FLAG-FLNA and 40 μl of FLAG-specific mAb M2 agarose (50% (v/v) slurry, Sigma) in binding buffer (25 mM Tris-HCl, 50 mM NaCl, 0.1% (w/v) Tween 20, 1 mM DTT, 10% sucrose, 5 mM MgCl₂, 1 mM EGTA, pH 7.4; 400 μl) for 2 h at 25°C . The beads were sedimented and washed three times with the binding buffer. Proteins bound to the beads were solubilized in SDS sample buffer and separated by 10.0% (w/v) SDS-PAGE followed by immunoblotting using mouse mAb against His conjugated with horseradish peroxidase (Sigma). For the peptide pull-down assay, a synthetic peptide of the β_7 -integrin cytoplasmic domain (Cys-⁷⁷¹KQDSNPLYKSAITTTINPR⁷⁸⁹) was immobilized on Sulfo-Link agarose beads (1 mg ml⁻¹) and mixed with increasing amounts of wild-type and mutant (del41, AA/DK, A2272D/A2274K) FLNA. Bound FLNA was detected by immunoblotting using mouse mAb to FLNA.

Reconstitution of actin-FLNA networks. Activity buffer¹⁰ (AB; 25 mM imidazole, 150 mM KCl, 5 mM MgATP, 0.2 mM CaCl₂, and 1 mM DTT, pH 7.4) and F-actin polymerization buffer³³ (FB; 20 mM Tris-HCl, 2 mM MgCl₂, 100 mM KCl, 0.2 mM DTT, 0.2 mM CaCl₂, 0.5 mM ATP, pH 7.5) were formulated as established previously. Shear cell samples consisted of 24 μM actin, 0.12 μM FLNA, $\times 1$ FB, 2 μM Alexa 546 phalloidin and either PA-GFP FilGAP or PA-GFP β_7 -integrin.

Shear-cell design and measurements. A P-780 (Physik Instrumente) piezo-motor, incorporated into a home-built microscope stage, was controlled using LABVIEW software (National Instruments). The sample component of the microscope stage had as its lower plate a stationary glass coverslip within a stainless steel frame and had a glass upper plate connected by a steel post to the piezo-motor. The gap between the lower and upper plates was 300 μm . A lateral shear was applied as illustrated in Supplementary Fig. 8. Strain is defined as the change of length divided by the original length. Thus, a 300- μm vertical distance that is deformed to 312 μm when sheared laterally by 85 μm is under an engineering strain of 1.04,

or 4%. The parameter γ is defined by the lateral shear divided by the sample thickness, yielding 85/300, or 0.28. Using a MATLAB-based Monte Carlo simulation of affine deformation, we calculated that a uniaxial shear strain of 0.28 causes the angle of an isotropic distribution of FLNA crosslinks initially at 90° to increase and decrease their opening angles symmetrically about 90° (on respective sides of the initially perpendicular intersection; Supplementary Fig. 8). Looking at the positive half of the distribution suggests that the weighted mean increase in opening angle is $\sim 6.1^{\circ}$, with a peak of $\sim 7.6^{\circ}$.

Myosin II system. To examine the effects of cytoskeleton-induced stress, and as a physiological technique complementary to external shear, we included 1 μM myosin II in the networks to generate contractile stress as illustrated in Supplementary Fig. 9 and Supplementary Movie 3. Individual myosin II molecules bind their tail regions together to form minifilaments, bipolar assemblies of 8–13 myosin molecules²¹. These minifilaments allow the otherwise non-processive myosin II to operate collectively with an increased duty cycle, binding the minifilaments to actin filaments long enough for filament sliding and network contraction to occur. At 150 mM KCl, approximately 8–13 myosin molecules associate into each minifilament²¹. The number of myosin minifilaments per actin filament, $N_{\text{mf/fil}}$, may be calculated from $N_{\text{mf/fil}} = [m]n_{\text{fil}}/[a]n_{\text{mf}}$, where $[m]$ is the molar concentration of myosin, n_{fil} is the number of actin monomers per actin filament, $[a]$ is the molar concentration of actin and n_{mf} is the number of myosin molecules per minifilament²². On the basis of an average actin filament length of $\sim 5 \mu\text{m}$ and there being 13 myosin molecules per minifilament, we estimate that approximately six minifilaments bind per actin filament. Repeating this estimation with FLNA instead of myosin, we estimate that there are five FLNA crosslinks per actin filament, on the basis of each crosslink being composed of two FLNA molecules. Thus, the density of myosin minifilaments per actin filament is approximately equal to that of FLNA crosslinks.

Samples were composed of 24 μM actin, 0.12 μM FLNA, 1 μM myosin II, $\times 1$ AB and 2 mM caged ATP, along with 2 μM Alexa 546 phalloidin and 50 nM PA-GFP FilGAP or 2 μM Alexa 488 phalloidin and 50 nM PA-mCherry β_7 -integrin.

Each sample was allowed to polymerize and consume available ATP over ~ 6 h, at which point myosin minifilaments had ceased contracting and were in a rigor state. FLAC measurements were then performed on the ATP-free unstressed network. Subsequently, the caged ATP (Sigma) was released by a 4-s exposure to a diffuse, 50-mW, 365-nm light-emitting diode (Prizmatix). Within 3 s, the network could be seen to become active, and it substantially homogenized within approximately 5 min. FLAC measurements were repeated in the active myosin-stressed network 10–20 min after ATP release, to quantify the strain-dependent binding activity.

Imaging and analysis. Fluorescence recovery after photobleaching (FRAP) provides an effective method to measure the diffusion and binding interactions of fluorescent proteins¹¹; however, the time resolution of FRAP is not adequate for fast transient events. To overcome this time limitation, we have developed FLAC, which takes advantage of the rapid photoactivation or conversion of novel PAFPs³¹ (Supplementary Figs 4 and 5). In FLAC, an initially non-fluorescent sample is locally pulsed with 405-nm light, activating PAFP fluorescence, and the time-dependent decay of the PAFP intensity reveals its diffusion and binding kinetics. An additional advantage is that FLAC may present more accurate binding/diffusion information than FRAP, as adequate photobleaching requires harsh conditions³⁴.

Time-dependent fluorescence images were acquired with a confocal microscope (TCS SP5, Leica) using a $\times 10$, 0.3 NA objective with images captured every 30–100 ms. PAFPs were activated using a 50-ms, ~ 5 -mW shuttered pulse from an external, 405-nm laser (Bestofferbuy) coupled into the confocal microscope using its non-descanned X1 port. A custom-built filter cube was installed with the filter holder rotated azimuthally by 90° to allow the 405-nm laser light entering the X1 port to be reflected upwards to the objective through a long-pass dichroic filter (Di01-R405-25x36, Semrock) and illuminate a 2- μm spot in the centre of the sample x - y plane.

Fluorescence in the photoactivation region was monitored in time and constitutes the principal data for the experiments. Data was collected before, during and after photoactivation and analysed in MATLAB. Fluorescence before activation represents the background signal, and was mainly a mixture of detection noise and background fluorescence. The background was defined as zero and subtracted from the data for each measurement. Data collected after activation was normalized to its maximum value, which was reached ~ 10 ms after photoactivation. Thus, fluorescence intensity data sets range from 1 at time zero after normalization and approach zero (background) at long times. Although a wide variety of fitting algorithms and procedures exist for FRAP, which in its analysis is mathematically similar to FLAC, we use a single exponential decay, $I(t) = ae^{-t/k} + c$, to fit our data. These fits are accurate, allow us to differentiate data quantitatively as simply as possible with a single time constant, k (the characteristic fluorescence decay time),

and do not invoke a variety of free parameters or models with a-priori assumptions¹¹. In each case, the fluorescence intensity decay is quantified and numerically compared using the same individual sample with and without strain to ensure consistency, eliminate artefacts or variability, and isolate mechanical strain as the single variable. In the case of single-step uniaxial strain presented in Fig. 2, data were fitted to $I(t) = ae^{-t/k} + 0.5e^{-t/(0.15s)} + c$ to provide a more accurate fit and compensate for the rapid diffusion of free fluorophore.

Photobleaching decays times were measured by photoactivating PAFP nonspecifically adsorbed to the glass surface, and were found to be in excess of 800 s, making them negligible on the timescale of FLAC experiments (<10 s), as shown in Supplementary Fig. 7. Photoconversion times were determined by acquiring

fluorescence intensity images of PAFP adsorbed to glass every 3 ms during a 50-ms, 405-nm activation flash, as shown in Supplementary Fig. 4.

31. Fernández-Suárez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nature Rev. Mol. Cell Biol.* **9**, 929–943 (2008).
32. van der Flier, A. *et al.* Different splice variants of filamin-B affect myogenesis, subcellular distribution, and determine binding to integrin β subunits. *J. Cell Biol.* **156**, 361–376 (2002).
33. Nakamura, F., Osborn, E., Janmey, P. A. & Stossel, T. P. Comparison of filamin A-induced cross-linking and Arp2/3 complex-mediated branching on the mechanics of actin filaments. *J. Biol. Chem.* **277**, 9148–9154 (2002).
34. Weissmann, C. *et al.* Microtubule binding and trapping at the tip of neurites regulate tau motion in living neurons. *Traffic* **10**, 1655–1668 (2009).

S-nitrosylation of NADPH oxidase regulates cell death in plant immunity

Byung-Wook Yun¹, Angela Feechan^{1†}, Minghui Yin¹, Noor B. B. Saidi^{1,2}, Thierry Le Bihan³, Manda Yu¹, John W. Moore¹, Jeong-Gu Kang¹, Eunjung Kwon¹, Steven H. Spoel¹, Jacqueline A. Pallas^{4†} & Gary J. Loake¹

Changes in redox status are a conspicuous feature of immune responses in a variety of eukaryotes^{1,2}, but the associated signalling mechanisms are not well understood. In plants, attempted microbial infection triggers the rapid synthesis of nitric oxide^{3,4} and a parallel accumulation of reactive oxygen intermediates, the latter generated by NADPH oxidases related to those responsible for the pathogen-activated respiratory burst in phagocytes⁵. Both nitric oxide and reactive oxygen intermediates have been implicated in controlling the hypersensitive response, a programmed execution of plant cells at sites of attempted infection^{3,5,6}. However, the molecular mechanisms that underpin their function and coordinate their synthesis are unknown. Here we show genetic evidence that increases in cysteine thiols modified using nitric oxide, termed S-nitrosothiols, facilitate the hypersensitive response in the absence of the cell death agonist salicylic acid and the synthesis of reactive oxygen intermediates. Surprisingly, when concentrations of S-nitrosothiols were high, nitric oxide function also governed a negative feedback loop limiting the hypersensitive response, mediated by S-nitrosylation of the NADPH oxidase, AtRBOHD, at Cys 890, abolishing its ability to synthesize reactive oxygen intermediates. Accordingly, mutation of Cys 890 compromised S-nitrosothiol-mediated control of AtRBOHD activity, perturbing the magnitude of cell death development. This cysteine is evolutionarily conserved and specifically S-nitrosylated in both

human and fly NADPH oxidase, suggesting that this mechanism may govern immune responses in both plants and animals.

Complex plants do not possess a nitric oxide synthase structurally related to those found in animals; nevertheless, a number of potential sources for pathogen-triggered nitric oxide synthesis have been described, including nitrate reductase and an arginine-dependent nitric-oxide-synthase-like activity⁷. S-nitrosylation, the addition of a nitric oxide moiety to a reactive cysteine thiol to form an S-nitrosothiol⁸ (SNO), is an important route for nitric oxide bioactivity. In *Arabidopsis*, an S-nitrosogluthione (GSNO) reductase (AtGSNOR1) governs both the concentrations of GSNO and, indirectly, protein SNOs⁸. We determined the temporal profile of SNO concentrations during the development of hypersensitive response in *atgsnor1-3* and *atgsnor1-1* plants, in which AtGSNOR1 activity is absent or increased, respectively⁸. Thus, SNO concentrations were anticipated to be higher in *atgsnor1-3* and lower in *atgsnor1-1* plants, relative to wild type.

SNO concentrations were also determined in the *NO overproducing 1* (*nox1*) mutant⁹. Such plants were challenged with the bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000 expressing either AvrB or AvrRps4 effector proteins, which are recognized by the resistance (*R*) gene products RPM1 and RPS4, respectively, each a prototypic member of a distinct *R* protein subclass^{10,11}. In each case, SNO concentrations increased over time in all the *Arabidopsis* lines tested (Fig. 1a, b), relative to *Pst* DC3000 controls (Supplementary Fig. 1). However,

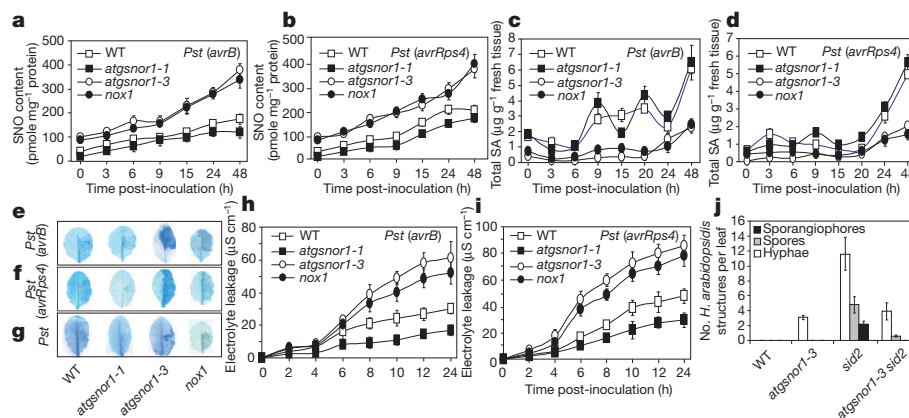


Figure 1 | SNOs positively regulate cell death by hypersensitive response. **a**, Profile of SNO accumulation following challenge with *Pst* DC3000 (*avrB*). **b**, SNO accumulation following attempted *Pst* DC3000 (*avrRps4*) infection. **c**, Total salicylic acid (SA) accumulation in response to attempted *Pst* DC3000 (*avrB*) colonization. **d**, Accumulation of salicylic acid following *Pst* DC3000 (*avrRps4*) challenge. **e**, Cell death development in the given *Arabidopsis* genotypes, triggered by 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 (*avrB*) at 15 h.p.i. and scored by trypan blue staining. **f**, Magnitude of cell death development in the stated plant lines following challenge with 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 (*avrRps4*) at 15 h.p.i.,

determined by trypan blue staining. **g**, Cell death development in the given *Arabidopsis* genotypes, triggered by 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 at 15 h.p.i. and scored by trypan blue staining. **h**, Extent of cell death development established by electrolyte leakage in the given *Arabidopsis* genotypes following challenge with *Pst* DC3000 (*avrB*) at 1×10^8 c.f.u. ml⁻¹. **i**, Quantification of cell death by electrolyte leakage in the stated plant lines in response to *Pst* DC3000 (*avrRps4*). **j**, Growth of *H. arabidopsidis* Emwa1 at 10 d post-inoculation in the given *Arabidopsis* genotypes. Data points represent mean \pm s.e. ($n = 3$). Unless stated otherwise, avirulent strains of *Pst* DC3000 were infiltrated at 1×10^6 c.f.u. ml⁻¹.

¹Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JH, UK. ²University of Putra, Department of Cell and Molecular Biology, Serdang 43400, Malaysia. ³University of Edinburgh, Centre for Systems Biology, Edinburgh EH9 3JR, UK. ⁴Trait Research, Syngenta, Jealott's Hill, Bracknell, Berkshire RG42 6EY, UK. [†]Present addresses: CSIRO Plant Industry, PO Box 350, Glen Osmond, South Australia 5064, Australia (A.F.); Bloomsbury Centre for Bioinformatics, University College London, Gower Street, London WC1E 6BT, UK (J.A.P.).

concentrations of these molecules were higher in *atgsnor1-3* and *nox1* plants than in wild type. Similar results were obtained when we used reporters of NO accumulation¹² (Supplementary Fig. 2a–d).

We also determined the profile of salicylic acid accumulation (salicylic acid is a cell death agonist¹³) in these lines. Total salicylic acid accumulation was diminished in *atgsnor1-3* and *nox1* plants relative to wild type (Fig. 1c, d and Supplementary Fig. 3a), as was free salicylic acid and salicylic acid β -glucoside (Supplementary Fig. 3b–e). Together, these results suggest that *atgsnor1-3* and *nox1* plants accrue markedly more SNOs over time during the development of hypersensitive response, and that the *atgsnor1-1* line accumulates significantly fewer. Further, salicylic acid concentrations are diminished in *atgsnor1-3* and *nox1* plants.

Next we assessed the development of hypersensitive response in these plants. Challenge with *Pst* DC3000 expressing either *avrB* or *avrRps4* revealed that this defence response was delayed in *atgsnor1-1* plants relative to wild type. In contrast, the development of hypersensitive response in *atgsnor1-3* and *nox1* plants was accelerated (Supplementary Fig. 4a, b). To determine the extent of cell death by hypersensitive response (CDHR), a smaller inoculum of *Pst* DC3000 strains was used and the resulting leaves were stained with trypan blue, which marks dead or dying plant cells⁵. Relative to wild type, *atgsnor1-3* and *nox1* plants showed a prominent increase in cell death, but this response was markedly reduced in the *atgsnor1-1* line (Fig. 1e–g). We corroborated these findings by quantifying cell-death-induced electrolyte leakage. Again, cell death was significantly greater in *atgsnor1-3* and *nox1* plants than in wild type, and there was a decrease in the *atgsnor1-1* line (Fig. 1h, i and Supplementary Fig. 5). To confirm and extend these findings, we studied the effect of high SNO concentrations mediated by *atgsnor1-3* on the hypersensitive response in the absence of SALICYLIC ACID INDUCTION DEFICIENT 2 (*SID2*) function, which is required for pathogen-triggered salicylic acid synthesis¹⁴. There was no significant difference in the extent of cell death development in the *atgsnor1-3 sid2* double mutant relative to the *atgsnor1-3* line (Supplementary Fig. 6). Collectively, these findings imply that in *atgsnor1-3* and *nox1* plants, the development of CDHR has accelerated kinetics and increased magnitude. However, in *atgsnor1-1* plants, cell death development is reduced. Thus, despite diminished salicylic acid concentrations, a greater concentration of SNO positively regulates the development of the hypersensitive response mediated by at least two distinct *R* gene subclasses.

CDHR does not seem to be required for limiting bacterial infection^{5,15}. Therefore, to identify a potential role for SNO-driven cell death in disease resistance, we challenged *atgsnor1-3* plants with the avirulent oomycete *Hyaloperonospora arabidopsidis* isolate Emwa1, which is recognized by *RPP4* (ref. 16). The death of challenged host cells has been proposed as a key resistance mechanism against oomycetes^{5,17}. As expected, cell death was more pronounced in *atgsnor1-3* plants, relative to wild type, in response to Emwa1 (Supplementary Fig. 7). In addition to its role in cell death, salicylic acid is also a key immune activator, and plants defective in its accumulation routinely show diminished defence responses¹⁴. Indeed, the *Arabidopsis sid2* mutant, in which pathogen-induced salicylic acid accumulation is reduced^{14,18}, was compromised in *RPP4*-mediated resistance against Emwa1 (Fig. 1j). By contrast, even though salicylic acid concentrations were equally reduced, relative to wild type, in *atgsnor1-3* plants (Supplementary Fig. 8a–c), Emwa1 failed to complete its life cycle in these plants in the same way it did in the resistant wild-type line (Fig. 1j). An *atgsnor1-3 sid2* double mutant also had increased resistance against Emwa1 relative to *sid2* plants (Fig. 1j). Moreover, although SNO and nitrite concentrations are higher in *atgsnor1-3* plants, in *sid* mutants they are comparable to wild type (Supplementary Fig. 9a, b). Therefore, cell death development mediated by increased SNO is sufficient to convey resistance against Emwa1 in the absence of salicylic acid accumulation and associated defence responses.

NO function is thought to be closely interconnected with that of reactive oxygen intermediates (ROIs) in cell death development^{3,6}. We therefore monitored ROI accumulation activated by distinct *R* proteins in *atgsnor1* and *nox1* mutants, as determined by 3',3'-diaminobenzidine (DAB) staining⁵. Relative to wild type, *atgsnor1-3* and *nox1* plants showed decreased pathogen-induced ROI accumulation, whereas mutant *atgsnor1-1* plants accumulated more ROIs even in the absence of pathogen challenge (Fig. 2a, b and Supplementary Fig. 10a, b). Hence, in addition to being autonomous of salicylic acid, increased SNO concentrations may also facilitate CDHR independently of DAB-detectable ROI accumulation. To explore this possibility further, we studied the effect of high SNO concentrations on the hypersensitive response in the absence of the NADPH-dependent oxidases AtRBOHD and AtRBOHF, which drive pathogen-induced ROI synthesis⁵. As expected, *atrbohD* and *atrbohF* single and double mutants showed decreased pathogen-induced CDHR than did wild-type plants, indicating that ROI synthesis is required for full development of the hypersensitive response (Fig. 2c and Supplementary Fig. 11). However, CDHR was not significantly different in *atgsnor1-3 atrbohD*, *atgsnor1-3 atrbohF* and *atgsnor1-3 atrbohD atrbohF* mutants than in *atgsnor1-3* plants (Fig. 2c and Supplementary Figs 11 and 12a), despite the reduced ROI accumulation in these lines (Supplementary Fig. 12b). These results suggest that high SNO concentrations can facilitate CDHR independently of ROI synthesis mediated by AtRBOHD or AtRBOHF.

Next we asked how SNO concentrations governed by AtGSNOR1 are able to manipulate ROI concentrations. AtRBOHD is responsible for virtually all the DAB-detectable ROIs induced by avirulent strains of *Pst* DC3000⁵. Thus, it is plausible that SNO concentrations regulate AtRBOHD function. However, basal or pathogen-triggered changes in SNO concentration did not influence AtRBOHD protein concentrations, as determined by monitoring a Myc–AtRBOHD fusion protein detected by a Myc-specific antibody in *atrbohD atgsnor1* double

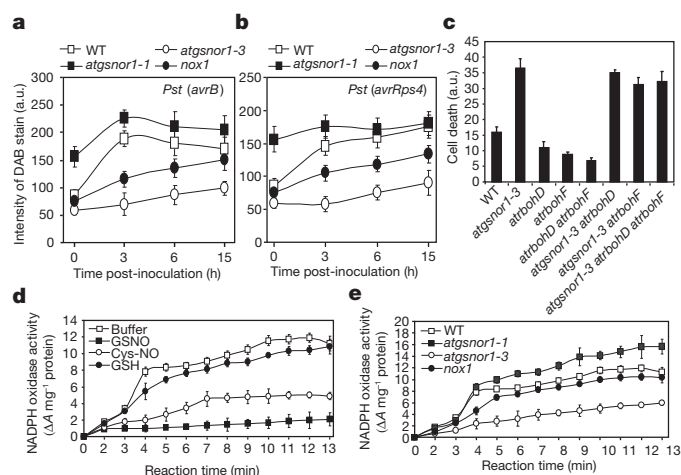


Figure 2 | Increased SNO concentrations blunt NADPH oxidase activity and reduce ROI accumulation. **a**, ROI accumulation determined by DAB staining in the given *Arabidopsis* genotypes following challenge with *Pst* DC3000 (*avrB*). a.u., arbitrary units. **b**, Accumulation of ROIs in the stated plant lines in response to attempted *Pst* DC3000 (*avrRps4*) infection. **c**, Cell death development in *atgsnor1* double and triple mutants in response to attempted *Pst* DC3000 (*avrB*) infection at 1×10^6 c.f.u. ml⁻¹ at 48 h.p.i. A Student's *t*-test comparing cell death in *atgsnor1-3* with that in *atgsnor1-3 atrbohF* plants, in a similar fashion to the other double and triple mutants, established that there was no statistically significant difference ($P = 0.6466$). **d**, NADPH oxidase activity in *Arabidopsis* following exposure to the given natural nitric oxide donors or related control treatments. ΔA at 480 nm, change in absorbance. **e**, NADPH oxidase activity in given *Arabidopsis* lines at 24 h.p.i. following challenge with *Pst* DC3000 (*avrB*). Data points represent mean \pm s.e. ($n = 3$). Avirulent strains of *Pst* DC3000 were infiltrated at 1×10^6 c.f.u. ml⁻¹ unless stated otherwise.

mutants (Supplementary Fig. 13). To explore whether SNOs could directly regulate NADPH oxidase activity, microsomal preparations from pathogen-challenged wild-type leaves were treated with the natural nitric oxide donor GSNO or *S*-nitroso-L-cysteine (Cys-NO), and NADPH oxidase activity was determined. Exposure to either GSNO or, to a lesser extent, Cys-NO significantly reduced the activity of this enzyme relative to the buffer control treatment (Fig. 2d). Furthermore, the absence of an effect following exposure to reduced glutathione (GSH) confirmed the specificity of this response. To determine the possible biological consequences of these findings, we measured the activity of this protein in *atgsnor1-3*, *atgsnor1-1*, *nox1* and wild-type leaves challenged with *Pst* DC3000 (*avrB*). NADPH oxidase activity was significantly reduced in *atgsnor1-3* and *nox1* plants that have high SNO concentrations (Fig. 2e). Collectively, these findings suggest that changes in SNO concentrations can modulate NADPH oxidase activity, implying that this protein might be regulated by *S*-nitrosylation.

To determine whether NADPH oxidase might be *S*-nitrosylated, recombinant protein was exposed to a range of either GSNO or Cys-NO concentrations typically used to score for *S*-nitrosylation *in vitro*^{19–21}, and was monitored for the formation of SNO–AtrBOHD by the biotin-switch method²². AtrBOHD was *S*-nitrosylated in a concentration-dependent fashion by either GSNO or Cys-NO (Fig. 3a, b). Furthermore, the addition of dithiothreitol strikingly reduced the concentration of SNO–AtrBOHD formation (Fig. 3b), consistent with the presence of a reversible thiol modification. AtrBOHD contains a number of cysteines that might serve as sites for this redox-based modification. The carboxy- and amino-terminal portions of this protein were therefore expressed, exposed to GSNO and subjected to biotin-switch analysis. Only the C-terminal portion of AtrBOHD was *S*-nitrosylated (Fig. 3c). The C-terminal portion of AtrBOHD has two cysteines: Cys 825 and Cys 890. These residues were therefore mutated either individually or in combination, and the resulting proteins were expressed, treated with GSNO and analysed by the biotin-switch method. The Cys890Ala mutation but not the Cys825Ala mutation abolished *S*-nitrosylation of AtrBOHD (Fig. 3d). Mass spectrometry analysis was also consistent with *S*-nitrosothiol formation at Cys 890 (Supplementary Fig. 14). Collectively, these findings imply that AtrBOHD is specifically *S*-nitrosylated *in vitro* on Cys 890.

Cys 890 is evolutionarily conserved, suggesting that NADPH oxidases from other eukaryotes might also be *S*-nitrosylated (Supplementary Fig. 15). We therefore exposed recombinant human and *Drosophila* RBOH proteins to either GSNO or Cys-NO. Both of these proteins were specifically *S*-nitrosylated (Supplementary Fig. 16a, b). The site of SNO formation was Cys 537 for human NOX2 (also known as CYBB) and Cys 1315 for *Drosophila* Nox (Supplementary Fig. 16c,

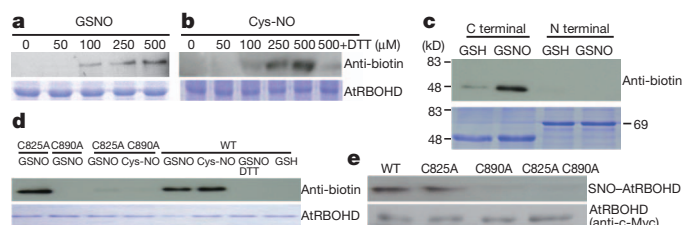


Figure 3 | *S*-nitrosylation of AtrBOHD. **a**, *S*-nitrosylation of recombinant AtrBOHD *in vitro* by the nitric oxide donor GSNO. **b**, Cys-NO *S*-nitrosylates recombinant AtrBOHD and this is reversible by treatment with dithiothreitol (DTT). **c**, The C terminus of AtrBOHD is *S*-nitrosylated by GSNO. **d**, *S*-nitrosylation analysis of wild-type and mutant AtrBOHD derivatives. **e**, *In vivo* *S*-nitrosylation of wild-type and mutant derivatives of AtrBOHD in the given plant genotypes 24 h after 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 (*avrB*) infiltration. AtrBOHD was detected by virtue of its c-Myc tag using antibodies raised against this tag. AtrBOH proteins in **a–d** were examined 20 min post-exposure to nitric oxide donors. All experiments were repeated at least twice, with similar results.

d, both of which corresponded to Cys 890 of AtrBOHD (Supplementary Fig. 15). Together, these data suggest that NADPH oxidases from at least two animals are specifically *S*-nitrosylated at this conserved cysteine, raising the possibility that this redox modification might regulate the activity of these enzymes in many other eukaryotes. Interestingly, an organizer protein, absent from plants, that interacts with animal NADPH oxidases may also be subject to *S*-nitrosylation in endothelial cells²³. Therefore, in mammals NADPH oxidase function might also be regulated indirectly by another reactive cysteine residue.

To determine whether AtrBOHD is *S*-nitrosylated *in vivo* during the hypersensitive response, transgenic *atrbohD* lines expressing either Myc-tagged wild-type AtrBOHD or mutant derivatives were challenged with *Pst* DC3000 (*avrB*). Subsequently, endogenous proteins were subjected to biotin-switch analysis and biotinylated proteins purified with streptavidin beads. These proteins were then immunoblotted with an anti-Myc antibody. Both wild-type AtrBOHD and the Cys825Ala mutant were *S*-nitrosylated, but the Cys890Ala mutant and the Cys825Ala Cys890Ala double mutant were not (Fig. 3e). Thus, AtrBOHD is specifically *S*-nitrosylated *in vivo* at Cys 890 during the plant defence response.

To understand whether *S*-nitrosylation of Cys 890 can modulate the activity of AtrBOHD, we first computationally modelled the structure of this protein. This indicated that Cys 890 is positioned closely behind the conserved Phe 921 and Phe 570 residues in AtrBOHD and NOX2, respectively. Similar to the homologous Tyr 247 in ferredoxin reductase²⁴, these residues are expected to have a significant role in binding flavin adenine dinucleotide (FAD). Accordingly, mutation of Phe 570 was reported to impair NOX2 function²⁵. The model further predicts that *S*-nitrosylation of AtrBOHD at Cys 890 may disrupt the side-chain position of Phe 921, impeding FAD binding (Supplementary Fig. 17a, b). To extend these findings, we determined the consequences of introducing an *S*-nitrosylated Cys 890 into our model. This disrupted the coplanar localization of Phe 921, thereby destabilizing or sterically ejecting FAD (Supplementary Fig. 18a, b). Consistent with these predictions, we found that prior GSNO exposure markedly reduced the binding of this cofactor to AtrBOHD. Conversely, GSNO did not diminish FAD binding in the Cys 890 AtrBOHD mutant (Fig. 4a).

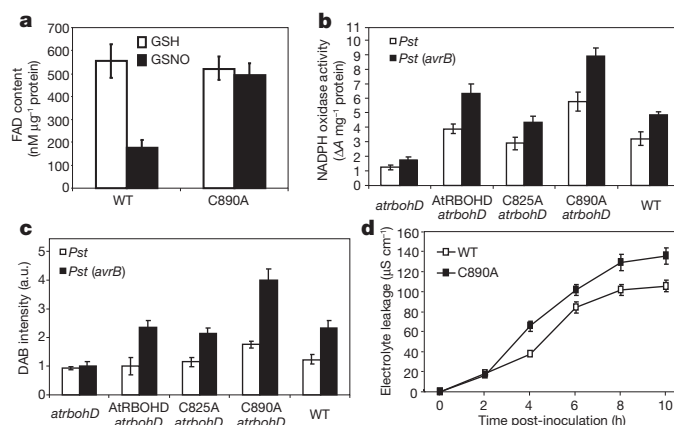


Figure 4 | The AtrBOHD Cys890Ala mutant shows increased activity during the defence response, amplifying ROI accumulation and cell death development. **a**, Quantification of FAD binding following prior GSNO or GSH treatment of wild-type AtrBOHD or the mutant Cys890Ala derivative. **b**, NADPH oxidase activity in plant extracts from the given *Arabidopsis* genotypes infiltrated with *Pst* DC3000 (white) or *Pst* DC3000 (*avrB*) (black) at 24 h.p.i. **c**, ROI accumulation determined by DAB staining in the leaves of given plant lines infiltrated with *Pst* DC3000 (white) and *Pst* DC3000 (*avrB*) (black) at 24 h.p.i. **d**, Extent of cell death development determined by electrolyte leakage in the given wild-type and mutant *Arabidopsis* lines following challenge with *Pst* DC3000 (*avrB*) infiltrated at 1×10^6 c.f.u. ml⁻¹. Data points represent mean \pm s.e. ($n = 3$). Unless stated otherwise, all strains of *Pst* DC3000 were infiltrated at 1×10^6 c.f.u. ml⁻¹.

Thus, S-nitrosylation of AtRBOHD may preclude FAD binding inhibiting the activity of this enzyme.

To explore the possible biological consequence of AtRBOHD SNO formation at Cys 890 and the resulting loss of FAD binding, we monitored NADPH oxidase activity in mutant *atrbohD* plants expressing either a wild-type AtRBOHD transgene or the Cys890Ala derivative. Pathogen-induced levels of leaf NADPH oxidase activity were significantly increased in the Cys890Ala mutant, relative to wild type (Fig. 4b). Furthermore, pathogen-induced ROI accumulation was greater in the Cys890Ala mutant line than in wild type (Fig. 4c). These data imply that S-nitrosylation of AtRBOHD at Cys 890 during the development of hypersensitive response blunts NADPH oxidase activity, resulting in decreased ROI generation. To determine the possible impact of the AtRBOHD Cys890Ala mutation on cell death, we challenged plants possessing this mutation with *Pst* DC3000 (*avrB*) and scored the extent of their response. There was a significant increase in cell death in the AtRBOHD Cys890Ala line relative to wild type (Fig. 4d). Thus, enhanced ROI production in the Cys890Ala mutant, which is not a target for S-nitrosylation, facilitates CDHR. This work was confirmed and extended by determining in parallel SNO accumulation, NADPH oxidase activity and cell death development in the AtRBOHD Cys890Ala mutant line, relative to wild type, following attempted pathogen ingress (Supplementary Fig. 19a–c). Together, this information implies that S-nitrosylation of AtRBOHD at Cys 890 promoted by increasing SNO concentrations serves to curb excessive cell death by blunting AtRBOHD-dependent ROI synthesis.

These data establish a molecular framework for SNO function during the hypersensitive response. We speculate that on attempted infection, total cellular SNOs governed by ATGSNOR1 and ROIs synthesized by AtRBOHD⁵, in combination with salicylic acid accumulation¹³, positively regulate the development of cell death. However, as SNO concentrations increase during the pathogen-triggered nitrosative burst, salicylic acid accumulation is reduced and further S-nitrosylation of AtRBOHD at Cys 890 decreases ROI synthesis. Collectively, this molecular dialogue may serve to limit cell death development during the hypersensitive response (Supplementary Fig. 20).

Our findings are reminiscent of those in animals, where increasing concentrations of total cellular SNOs drive apoptosis in a variety of tissues¹². Conversely, S-nitrosylation of the pro-apoptotic regulators nuclear factor κ B²⁶ and caspase cysteine protease²⁷ function to constrain cell death development. Furthermore, SNO formation at a conserved cysteine in NADPH oxidases from plants, humans and flies implies that this novel pro-survival mechanism might operate during immune function throughout such complex eukaryotes.

METHODS SUMMARY

Pathogen inoculations and small-molecule measurements. A variety of *Pst* DC3000 strains and *H. arabidopsidis* isolates were used for pathogen inoculations, as indicated. Trypan blue and DAB staining were conducted for detection of cell death and ROI accumulation, respectively²⁸. Salicylic acid was determined using a mini-scale procedure²⁹ based on high-pressure liquid chromatography. SNO contents were analysed using a chemiluminescence-based procedure³⁰. Diaminofluorescein diacetate imaging was applied for the detection of nitric oxide. Electrolyte leakage was measured by using a DiST WP conductivity meter to quantify cell death.

Gene constructs and transgenic plants. For cloning of NADPH oxidase genes and site-directed mutagenesis, the pGEX4T-1 vector and the QuickChange II Site-Directed Mutagenesis Kit (Stratagene) were used, respectively. Transgenes were generated by using the pUNI51 cloning vector and the pKYLX-myc9-loxp binary vector, and subsequently were introduced into *Arabidopsis* plants. See Methods for full details.

Biochemistry and computational modelling. NADPH oxidase activity was determined by using epinephrine and NADPH as substrates. *In vitro* and *in vivo* S-nitrosylation assays used either an anti-biotin or an anti-Myc9 antibody, respectively. Mass spectrometry was done using a capillary gas high-pressure liquid chromatography tandem mass spectrometry analysis system. For computational

modelling of NOX2 and AtRBOHD, we used PyMol, DeepView, SwissModel and Jpred as described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 May 2010; accepted 4 August 2011.

Published online 2 October; corrected 13 October 2011 (see full-text HTML version for details).

- MacMicking, J. D. *et al.* Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc. Natl Acad. Sci. USA* **94**, 5243–5248 (1997).
- Tada, Y. *et al.* Plant immunity requires conformational changes of NPR1 via S-nitrosylation and thioredoxins. *Science* **321**, 952–956 (2008).
- Delledonne, M., Xia, Y., Dixon, R. A. & Lamb, C. J. Nitric oxide functions as signal in plant disease resistance. *Nature* **394**, 585–588 (1998).
- Durner, J., Wendehenne, D. & Klessig, D. F. Defense gene induction in tobacco by nitric oxide, cyclic GMP and cyclic ADP ribose. *Proc. Natl Acad. Sci. USA* **95**, 10328–10333 (1998).
- Torres, M. A., Dangel, J. L. & Jones, J. D. *Arabidopsis* gp91^{phox} homologues *AtRbohD* and *AtRbohF* are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc. Natl Acad. Sci. USA* **99**, 517–522 (2002).
- Delledonne, M., Zeier, J., Marocco, A. & Lamb, C. Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response. *Proc. Natl Acad. Sci. USA* **98**, 13454–13459 (2001).
- Leitner, M., Vandelle, E., Gaupels, F., Bellini, D. & Delledonne, M. NO signals in the haze: nitric oxide signalling in plant defence. *Curr. Opin. Plant Biol.* **12**, 451–458 (2009).
- Feechan, A. *et al.* A central role for S-nitrosothiols in plant disease resistance. *Proc. Natl Acad. Sci. USA* **102**, 8054–8059 (2005).
- He, Y. *et al.* Nitric oxide represses the *Arabidopsis* floral transition. *Science* **305**, 1968–1971 (2004).
- Grant, M. R. *et al.* Structure of the *Arabidopsis* RPM1 gene enabling dual specificity disease resistance. *Science* **269**, 843–846 (1995).
- Gassmann, W., Hinsch, M. E. & Staskawicz, B. J. The *Arabidopsis* RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease resistance genes. *Plant J.* **20**, 265–277 (1999).
- Liu, L. *et al.* Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**, 617–628 (2004).
- Shirasu, K., Nakajima, H., Rajasekhar, V. K., Dixon, R. A. & Lamb, C. J. Salicylic acid potentiates an agonist-dependent gain control that amplifies pathogen signals in the activation of defense mechanisms. *Plant Cell* **9**, 261–270 (1997).
- Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**, 562–565 (2001).
- Yu, I. C., Parker, J. & Bent, A. F. Gene-for-gene disease resistance without the hypersensitive response in *Arabidopsis* *dnd1* mutant. *Proc. Natl Acad. Sci. USA* **95**, 7819–7824 (1998).
- Holub, E. B., Beynon, J. L. & Crute, I. R. Phenotypic and genotypic characterization of interactions between isolates of *Peronospora parasitica* and accessions of *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* **7**, 223–239 (1994).
- Keller, H. *et al.* Pathogen-induced elicitor production in transgenic tobacco generates a hypersensitive response and nonspecific disease resistance. *Plant Cell* **11**, 223–236 (1999).
- Nawrath, C. & Metraux, J. P. Salicylic acid induction-deficient mutants of *Arabidopsis* express *PR-2* and *PR-5* and accumulate high levels of camalexin after pathogen inoculation. *Plant Cell* **11**, 1393–1404 (1999).
- Wang, Y.-Q. *et al.* S-nitrosylation of ATSBP3 antagonises the expression of plant immunity. *J. Biol. Chem.* **284**, 2131–2137 (2009).
- Romero-Puertas, M. C. *et al.* S-nitrosylation of peroxiredoxin II E promotes peroxynitrite-mediated tyrosine nitration. *Plant Cell* **19**, 4120–4130 (2007).
- Lindemayr, C., Sell, S., Müller, B., Leister, D. & Durner, J. Redox regulation of the NPR1-TGA1 system of *Arabidopsis thaliana* by nitric oxide. *Plant Cell* **22**, 2894–2907 (2010).
- Jaffrey, S. R., Erdjument-Bromage, H., Ferris, C. D., Tempst, P. & Snyder, S. H. Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nature Cell Biol.* **3**, 193–197 (2001).
- Selamidis, S., Dusting, G. J., Peshavariya, H., Kemp-Harper, B. K. & Drummond, G. R. Nitric oxide suppresses NADPH oxidase-dependent superoxide production by S-nitrosylation in human endothelial cells. *Cardiovasc. Res.* **75**, 349–358 (2007).
- Ingelman, M., Bianchi, V. & Eklund, H. The three-dimensional structure of flavodoxin reductase from *Escherichia coli* at 1.7 Å resolution. *J. Mol. Biol.* **268**, 147–157 (1997).
- Zhen, L., Yu, L. & Dinauer, M. C. Probing the role of the carboxyl terminus of the gp91^{phox} subunit of neutrophil flavocytochrome b₅₅₈ using site-directed mutagenesis. *J. Biol. Chem.* **273**, 6575–6581 (1998).
- Matthews, J. R. *et al.* Inhibition of NF- κ B DNA binding by nitric oxide. *Nucleic Acids Res.* **24**, 2236–2242 (1996).
- Mannick, J. B. *et al.* Fas-induced caspase denitrosylation. *Science* **284**, 651–654 (1999).
- Yun, B.-W. *et al.* Loss of actin cytoskeletal function and EDS1 activity, in combination, severely compromises non-host resistance in *Arabidopsis* against wheat powdery mildew. *Plant J.* **34**, 768–777 (2003).
- Aboul-Soud, M. A. M., Cook, K. & Loake, G. J. Measurement of salicylic acid by a high-performance liquid chromatography procedure based on ion-exchange. *Chromatographia* **59**, 129–133 (2004).

30. Liu, L. *et al.* Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**, 617–628 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to acknowledge R. Innes and W. Gassmann for *Pst* DC3000 strains expressing either *avrB* or *avrRps4*, respectively. *Arabidopsis* transfer DNA insertion mutants were obtained from SAIL (Syngenta) populations. We thank M. Tör for the *H. arabidopsidis* isolate Emwa1, and K. Kanchanawatee for the *Drosophila* cDNA clone and associated mutant. A.F. was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) CASE studentship. B.-W.Y. and E.K. were funded by BBSRC grant BB/D011809/1 awarded to G.J.L. J.W.M. received a grant from the Physical Sciences Research Council (EPSRC). M. Yu was the recipient of a Darwin Trust Scholarship. T.L.B. was supported by BBSRC and EPSRC grant BB/D019621/1.

N.B.B.S. and M. Yin were funded by a Ministry of Education Malaysia scholarship and a Torrance Scholarship, respectively.

Author Contributions G.J.L. designed the research and wrote the paper, and with B.-W.Y. planned experiments and analyses. B.-W.Y. conducted the majority of experiments. A.F., M. Yin, N.B.B.S., J.-G.K., J.W.M., M. Yu, E.K. and T.L.B. conducted experiments. S.H.S. generated and interrogated three-dimensional models. J.A.P. was the industrial supervisor of A.F. All authors, especially B.-W.Y. and S.S., commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.L. (gloake@ed.ac.uk).

METHODS

Histochemical analysis and small-molecule measurements. Cells committed to die were visualized with lactophenol-trypan blue as described previously²⁸. Peroxides were stained with DAB²⁸. Catalase effectively eliminated DAB staining. H₂O₂-dependent DAB staining was observed in all the *Arabidopsis* lines described in this work. For quantification of either cell death or H₂O₂ accumulation, the extent of either trypan blue or DAB staining, respectively, was determined in six leaves per line in arbitrary units, for each time point and from three independent experiments, using the 'saturation' function in PaintShop Pro 8 (Corel).

Salicylic acid and salicylic acid β -glucoside concentrations were determined using a mini-scale procedure²⁹ based on high-pressure liquid chromatography. For the identification of SNOs, tissue extracts were generated at given times after pathogen inoculation and these samples were analysed by a chemiluminescence-based procedure (Sievers nitric oxide analyser) as described previously³⁰.

For diaminofluorescein diacetate imaging, rosette leaves were incubated for 15 min in a solution of 15 μ M diaminofluorescein diacetate (Alexis) containing 5 mM MES KOH, pH 5.7, 0.25 mM KCl and 1 mM CaCl₂, and were then washed for 5 min (ref. 31). Fluorescent signals were detected using a Bio-Rad Radiance 2100 confocal microscope (Nikon Eclipse, TE2000-U). The dye was excited at 488 nm, and images were collected at emission wavelengths of 500–530 nm. Green fluorescence-specific intensities were quantified using IMAGEJ (version 1.45h).

Plant material and pathogen inoculations. *Arabidopsis* accession Col-0 and cognate mutants were grown under 16 h of light at 22 °C and 8 h of darkness at 18 °C. *Pst* DC3000 was grown, maintained and inoculated as described previously³². *Pst* DC3000 strains were inoculated at the concentrations stated. *H. arabidopsidis* infections were as described previously¹⁶.

Electrolyte leakage. The protocol for electrolyte leakage was adapted from a previously described method³³. Four-week-old plants were injected with bacteria in 10 mM MgCl₂. Ten minutes after injection, 5.0-mm-diameter leaf disks were collected from the injected area and washed extensively with water for 10 min, and then ten discs were placed in a Petri dish with 6 ml of water. Conductivity measurements (three replicates for each treatment) were taken over time by using a DiST WP conductivity meter (HANNA Instruments). The units of this measurement are microsvdbergs per centimetre, where the distance refers to that between the electrodes.

Cloning of NADPH oxidase genes and site-directed mutagenesis. The primers used to clone the N terminus (from K2 to N357; 356 amino acids) and the C terminus (from K756 to F921; 166 amino acids) of AtRBOHD, the C terminus (from N430 to E540; 111 amino acids) of *Homo sapiens* NOX2 (gi_163854302) and the C terminus (from I1198 to E1338; 141 amino acids) of *Drosophila melanogaster* Nox (gi_161077139) are as follows. AtRBOHD N terminus: 5'-AAGGATCCAAAATGAGACGAGGCAATTCA-3' (forward); 5'-TTCTAG TTGCTCTCTTTGCCGGTCT-3' (reverse). AtRBOHD C terminus: 5'-AAG GATCCAAGGACATCATCAACAACATG-3' (forward); 5'-TTCTAGAAGTT CTCTTTGTGGAA-3' (reverse). NOX2 C terminus: 5'-GTAATGGATCCAAC GCCACCAATCT-3' (forward); 5'-TATGCTCTCGAGTCATTCAAGTCCACA GA-3' (reverse). Nox C terminus: 5'-GAAGAGCAAAAAGCGGAGTC-3' (forward); 5'-GGATTGGCTTTCGTAAGGA-3' (reverse). The amplified PCR products were cloned into pGEX4T-1 vector at the sites of BamHI/EcoRI, BamHI/XhoI and EcoRI/XhoI, respectively.

Site-directed mutagenesis was carried out with QuickChange II Site-Directed Mutagenesis Kit (Stratagene). All procedures followed the manufacturer's manual and specific primers used for AtRBOHD, NOX2 and Nox mutations are as follows. AtRBOHD C825A: 5'-GAGCTTCACAATTATGCCACGAGTGTGTACGA-3' (forward); 5'-TCGTACACACTCGTGGCATAATTGTGAAGCTC-3' (reverse). AtRBOHD C890A: 5'-ATAGGAGTCTTCTACGCTGGAATGCCAGGAAT-3' (forward); 5'-ATTCCTGGCATTCAGCTAGAAGACTCCTAT-3' (reverse). NOX2 C537A: 5'-ATAGGAGTTTCTCTGCCGGACCTGAATGACTC-3' (forward); 5'-GAGTCATTCAAGTCCGGCGAGGAAAACCTCTAT-3' (reverse). Nox C1315A: 5'-GTCACCGTCTTCTACGCCGGCCACCACAGTTG-3' (forward); 5'-CAACTGTGGTGGCGCCGCTAGAAGACGGTGAC-3' (reverse). Underlining denotes the codon of alanine.

Transgenic plant materials. The *atrbohD* line was complemented with a wild-type copy of AtRBOHD and also transformed with C825A, C890A or the double mutant. Briefly, a wild-type copy and a mutant copy of AtRBOHD were amplified by using forward primer 5'-CGGAATTCGGATGAAAATGAGACGAGGCAA TTCA-3' and reverse primer 5'-CGGGATCCTAGAAGTTCTCTTTGTGGA AGTC-3', and cloned into pUNI51 vector (EcoRI/BamHI). Subsequently, the pUNI51 constructs harbouring AtRBOHD were recombined with pKYLX-myc9-loxP binary vector by a Cre recombinase³⁴. The resulting constructs were introduced into *Agrobacterium* strain GV3101 and subsequently transformed into *Arabidopsis* plants. Transgenic T1 plants were identified by kanamycin selection.

NADPH oxidase biochemistry. NADPH oxidase activity was measured as described previously³⁵. Briefly, 1 g of leaf tissue was ground in liquid nitrogen and dissolved in extraction buffer (0.25 M sucrose, 50 mM HEPES, pH 7.2, 3 mM EDTA, 1 mM dithiothreitol, 0.6% PVP, 3.6 mM L-cysteine, 0.1 mM MgCl₂ and protease inhibitor tablet (Roche)). The crude extract was centrifuged at 10,000g for 30 min and the resulting supernatant was ultracentrifuged at 203,000g for 1 h. The resulting pellet was resuspended in extraction buffer and used as the membrane fraction to measure NADPH oxidase activity spectrophotometrically at 480 nm using epinephrine and NADPH as substrates.

Expression and purification of recombinant proteins. Recombinant proteins were expressed in *Escherichia coli* strain BL21 (DE3) by adding 0.3 mM IPTG with a 6-h incubation. The GST-tagged proteins were purified using a MagneGST Protein Purification System (Promega). All procedures followed the manufacturer's manual.

In vitro and vivo S-nitrosylation assays. Recombinant proteins were *in vitro* S-nitrosylated with the stated concentration of the given nitric oxide donor in 500 μ l volumes for 20 min in darkness. Donors were removed using Micro Bio-Spin P6 columns (BioRad) and the resulting proteins were subjected to the biotin-switch technique²² by western blot assay using anti-biotin antibody (New England Biolab). For *in vivo* assay, *Arabidopsis* plants were inoculated with *Pst* DC3000 (*avrB*) at 10⁷ c.f.u. ml⁻¹ and an anti-Myc9 antibody (Sigma) was used.

Mass spectrometry. All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Acetonitrile and water for liquid chromatography tandem mass spectrometry and sample preparation were of high-pressure liquid chromatography quality (Fisher). Formic acid was Suprapure 98–100% (Merck) and trifluoroacetic acid was 99% purity sequencing grade. Sequencing-grade modified porcine trypsin was purchased from Promega and GluC from Worthington (Lorne Lab). All high-pressure liquid chromatography mass spectrometry (LC-MS) connector fittings were from Upchurch Scientific or Valco (Hichrom and RESTEK).

As S-nitrosothiol formation at Cys 890 of AtRBOHD was found to be relatively labile *in vitro*, we used a well established method³⁶ that utilizes iodoacetamide to form a carbamidomethyl ion at sites of S-nitrosothiol formation, which are not blocked by treatment with methyl methanethiosulphonate (MMTS). This analysis revealed striking S-nitrosylation at Cys 890 of AtRBOHD but the complete absence of S-nitrosothiol formation at Cys 825 following treatment with a nitric oxide donor.

The GST-AtRBOHD C terminal was expressed in *E. coli* BL21 and purified using glutathione Sepharose 4B (GE Healthcare) in native condition. The purified protein solution was desalted using a Zeba Desalt spin column (Thermo Scientific). The desalted protein was treated with or without Cys-NO (final concentration, 0.5 mM) in HENS buffer for 20 min at room temperature (25 °C). Cys-NO was then removed using a Zeba desalt column and the free cysteines were blocked by MMTS in HENS buffer with 2.5% SDS for 20 min at 50 °C. The treated protein was precipitated by acetone and resuspended in HENS buffer with 1% SDS. Sodium ascorbate (10 mM) and iodoacetamide (50 mM) were added to remove S-NO bonds and for protein acylation. The proteins were separated by SDS-PAGE and the protein gel was excised, cleaned and digested with GluC and Trypsin at 37 °C for 16 h. The digested peptides were blocked first with MMTS and analysed by LC-MS.

Capillary gas high-pressure liquid chromatography tandem mass spectrometry (MSMS) analysis was performed on an online system consisting of a micropump (1200 binary HPLC system, Agilent) coupled to a hybrid LTQ-Orbitrap XL instrument (Thermo-Fisher). The LTQ was controlled through XCALIBUR 2.0.7. Samples were reconstituted in 10 μ l loading buffer before injection and analysed on a 1-h gradient for data-dependent analysis.

MSMS data were searched using MASCOT, versions 2.2 and 2.3 (Matrix Science Ltd), against a small database comprising the most common contaminant and various constructs. Variable methionine oxidation and cysteine methylthiol and carbamidomethylation were considered in all analyses. The precursor mass tolerance was set to 7 p.p.m. and the MSMS tolerance was set to 0.4 AMU. Fragmentation patterns were confirmed using ProteinProspector (<http://prospector.ucsf.edu>). Label-free quantitation was performed using PROGENESIS (Nonlinear Dynamics). For label-free quantitation, the number of 'features' (that is, signal at a specific retention time and *m/z*) was reduced to only MSMS peaks with a charge of 2, 3 or 4+ and only the five most intense MSMS spectra per feature²⁷ were kept. Sets of multicharged ions (2+, 3+, 4+) were extracted from each LC-MS run.

FAD-binding activity assay. FAD-binding activity was measured as described previously³⁷. Briefly, proteins purified using a MagneGST Protein Purification System (Promega) were incubated with either GSNO or GSH for 20 min in the dark, and the excess GSNO or GSH was removed using a Bio-Spin6 column (Bio-Rad). The resulting compounds were further incubated with FAD (250 μ M) for 30 min in the dark. Unbound FAD was removed using a Bio-Spin6 column, and

the FAD content was determined by boiling the resulting protein samples for 5 min in the dark, followed by centrifugation at 14,000g for 10 min to remove coagulated protein. The absorbance of the released FAD was measured at 450 nm.

Computational modelling of NOX2 and AtRBOHD. The C-terminal 287 amino acids of NOX2 were used for identifying structural homologues with Phyre³⁸. This identified PDB crystal structure 1FDR, encoding an *E. coli* flavodoxin reductase, as a potential homologue. Structural alignments were optimized using DeepView and submitted for threading over 1FDR at the SwissModel server³⁹. The resulting computational model contained several loop regions unique to NOX2. These regions were analysed for secondary structure using Jpred⁴⁰ and the DeepView loop library. The computational model of the AtRBOHD C terminus was built by threading over both 1FDR and NOX2.

The model indicates that Cys 890 is positioned closely behind the conserved Phe 921 and Phe 570 residues in AtRBOHD and NOX2, respectively. Similar to the homologous Tyr247 in ferredoxin reductase²⁴, these residues are expected to have a significant role in FAD binding. Accordingly, mutation of Phe 570 has been reported to impair NOX2 function²⁵. The model further predicts that S-nitrosylation of AtRBOHD at Cys 890 may disrupt the side-chain position of Phe 921, impeding FAD binding.

31. Foissner, I., Wendehenne, D., Langebartels, C. & Durner, J. *In vivo* imaging of an elicitor-induced nitric oxide burst in tobacco. *Plant J.* **23**, 817–824 (2000).

32. Whalen, M. C., Innes, R. W., Bent, A. F. & Staskawicz, B. J. Identification of *Pseudomonas syringae* pathogens of *Arabidopsis* and a bacterial locus determining avirulence on both *Arabidopsis* and soybean. *Plant Cell* **3**, 49–59 (1991).
33. Dellagi, A., Brisset, M.-N., Jean-Pierre Paulin, J.-P. & Expert, D. Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity. *Mol. Plant Microbe Interact.* **11**, 734–742 (1998).
34. Liu, Q., Li, M., Leibham, D., Cortez, D. & Elledge, S. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**, 1300–1309 (1998).
35. Sagi, M. & Fluhr, R. Superoxide production by plant homologues of the gp91^{phox} NADPH oxidase. Modulation of activity by calcium and by tobacco mosaic virus infection. *Plant Physiol.* **126**, 1281–1290 (2001).
36. Chen, Y. Y., Huang, Y. F., Khoo, K. H. & Meng, T. C. Mass spectrometry-based analyses for identifying and characterizing S-nitrosylation of protein tyrosine phosphatases. *Methods* **42**, 243–249 (2007).
37. Shen, A. L. & Kasper, C. B. Differential contribution of NADPH-cytochrome P450 oxidoreductase FAD binding site residues to flavin binding and catalysis. *J. Biol. Chem.* **275**, 41087–41091 (2000).
38. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
39. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
40. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201 (2008).

CAREERS

UNITED STATES Postdoc programme offers close-up view of industry life **p.277**

EUROPE Funding doesn't cover expenses for some doctoral students **p.277**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



Rookie review

What the novice peer reviewer needs to know before combing through a submission.

BY VIRGINIA GEWIN

Claudio Casola had no idea that journal editors had consistently rated his manuscript reviews highly. Then he received an award from Amsterdam-based publisher Elsevier for his “exceptional contribution to the quality of the journal *Gene*”. Casola is meticulous; he goes through every figure and table of a manuscript, tactfully suggests improvements and always meets his deadlines. He attributes his expertise to practice, publishing experience and immersion in the literature.

Casola, a postdoc in evolutionary genetics at Indiana University in Bloomington, says that his first review, in 2006, was typical of rookie referees. He spent more than 10 hours on the manuscript, poring over the details and asking faculty members for advice. After reviewing more than two dozen papers in the past five years, he has been able to cut the process down to three hours, quickly assessing the originality and merit of a paper. “Reviewing manuscripts

makes me feel like I’m a fully fledged member of the scientific community,” says Casola.

“Young scientists should get involved in the process as they start building their careers, particularly since reviewers are harder and harder to find,” says Bart Wacek, an executive publisher in charge of Elsevier’s genetics portfolio based in Cambridge, Massachusetts. Young

reviewers are certainly sought after. “The best referees are postdocs,” says Leslie Sage, a senior physical-sciences editor at *Nature* in Washington DC. “They are at the top of their game, well versed in the literature and politically naive enough to tell the truth.”

Yet Casola, like many graduate and postdoctoral students, was never taught how to review a manuscript; most peer reviewers learn journals’ needs and the reviewer’s role only through trial and error. Editors’ expectations differ according to their fields, but most agree that simply writing thorough, respectful and helpful reviews is the best way for early-career scientists to find their footing and avoid mistakes.

HELPFUL HINTS

The questions to ask

- Is this an innovative approach?
- How do the findings relate to other advances in the field?
- Are the methods appropriate to the scientific question being addressed?
- Does the paper adequately cite all the relevant literature?
- Are the data valid and do they support the conclusions? **V.G.**

PEER-REVIEW BLUEPRINT

When reading a paper, good reviewers will not only dig into the fine details of the research, but also assess the article’s contribution to the big picture of science. They should keep in mind what editors want most from them: detailed comments that justify their criticisms and are returned in good time. ►

BELLE MELLOR

► As online publication increases and new periodicals proliferate, journals are struggling to decrease time to publication. Many, including *Gene*, track reviewer performance and monitor turnaround time; reviewers who consistently return reviews after the deadline probably won't be called on again. "If you don't have the expertise or time to do a good review, decline right away," says Mark Spigt, a public-health researcher at Maastricht University in the Netherlands. Last year, Spigt wrote a set of detailed guidelines for reviewers, sharing insight gleaned from his own frustrations early in his career (M. Spigt and I. C. W. Arts *J. Clin. Epidemiol.* **63**, 1385–1390; 2010). "If you commit to the review, block out a set amount of time on the calendar and complete it in that window," he says.

Editors and seasoned reviewers recommend that referees start the review with a summary of the paper's findings and points of interest, to show that they have read and understood the article. They should also include an impact assessment, detailing whether the paper covers a huge or incremental advance, says Bernd Pulverer, chief editor of the *EMBO Journal* in Heidelberg, Germany. Spigt lists his major comments, then gives a more detailed, thorough account of the research. He addresses central issues such as whether the research question is relevant to the field, the methods are appropriate and the data support the conclusions (see 'The questions to ask').

Reviewers, who tend to be well acquainted with the literature, can best assist editors by comparing the manuscript with other advances in the field to ensure that it is saying something new and that no pertinent citations have been omitted — a task that is often overlooked. "In my personal experience, roughly 20% of authors are hiding relevant work in order to make the paper seem more novel," says Sage.

Manuscript editors appreciate informed assessments of the paper's findings and data presentation. "I like to see technical details — thoughts on the validity of the work, any concerns about technical issues, or potential improvement to the paper," says Daniel Kulp, director of the editorial division at the American Physical Society in Ridge, New York, which publishes *Physical Review Letters*, among other journals. The single most helpful thing a reviewer can do, editors say, is to substantiate critical comments with supporting evidence. "What's really frustrating is when a



"If you don't have the expertise or time to do a good review, decline right away."

Mark Spigt

INSIDE INFORMATION

What reviewers should know about reviewing

There are many common misconceptions about the review process. Here are some tips on the reviewer's role, what to do and what to avoid.

- Referees don't accept or reject papers; their comments guide the editor's decision.
- Reviews should not tear papers apart; it is better to offer constructive ideas on how to improve them.
- Reviewers should avoid concentrating on what the study could show in principle. The focus should be on what it actually shows.
- The stature of the scientist being reviewed

is unimportant. Referees should be neither hypercritical nor obsequious.

- Judging the science properly means setting aside one's own scientific biases or predilections.
- The reviewer should aim to improve the paper, not make someone else's work their own.
- Reviewers should not waste too much time trying to improve the writing.
- The journal editor will be able to clarify any uncertainties about conflicts of interest or the level of scrutiny expected. **V.G.**

reviewer simply says 'This is not good enough,' yet offers no constructive comments to remedy the problem," says Pulverer.

Some editors want to see thorough scrutiny of the data, with spot-checks of how uncertainties and outliers are addressed, for example. The level of checking required varies, so novice reviewers should ask editors what they look for. Some editors send cover letters detailing their expectations, but others don't bother.

Sage says it is not uncommon for his reviewers to download raw data to check that they are robust — especially if they are available from a public archive. He doesn't expect that much for every paper, but says that it can be helpful for a reviewer to take the initiative.

DETAILED ANALYSIS

Some fields and journals encourage extreme thoroughness. "We don't simply accept or reject; we have a detailed editing process that puts [the journal] on the high end of the involvement spectrum," says David Schimel, editor-in-chief of *Ecological Applications* and a senior scientist at the National Center for Atmospheric Research in Boulder, Colorado. Reviewers for *Ecological Applications* often submit a several-page report in which they not only evaluate methodology in considerable detail, check calculations and confirm whether the conclusions are supported, but also make clarifying comments on a line-by-line basis, he says. However, editors at other journals prefer reviewers not to spend their time on that level of detail, fearing that they will burn out.

Astronomy journals are generally comfortable with papers being revised several times, says Chris Sneden, an astronomer at the University of Texas at Austin and editor of *The Astrophysical Journal Letters*. "It's rare, but a paper can go through five or six review rounds if it starts out as a disaster," he says. "But the sociology of the field is happy with a lot of back and forth with the author during the process." This is less common for biology journals, which tend to have many more manuscripts moving through the

publication process at any one time.

Perhaps the biggest misconception about reviewing is that referees haven't done their job if they can't come up with enough criticisms. Overzealous reviewers may even allow criticism to devolve into cattiness in their eagerness not to be seen as a pushover. Postdocs are particularly prone to such errors, say some manuscript editors — perhaps because they are subconsciously trying to prove their own worth, or are used to journal-club sessions designed to pick papers apart. A lot of new reviewers simply can't write a review that essentially says 'this is a great paper' — which is perfectly valid, as long as reviewers summarize the key points and express in detail why the techniques and conclusions are sound," says Kulp. Hemai Parthasarathy, a former manuscript editor at *Nature* and *PLoS Biology* and now a senior partner at Torch Communications in San Francisco, California, advises that reviewers "start from a place of respect", and recognize that the author put a lot of work into the manuscript. "

Another mistake is to ask for lots of extra lab work (see 'What reviewers should know about reviewing'). "There are always more experiments that can be done to nail something down, and reviewers sometimes fall into the trap of asking for them all," says Parthasarathy. Novice reviewers have to learn to gauge when such requests are justified. "One of the trickiest things for junior people to learn is how meaty a manuscript needs to be," says Michael Veeman, a molecular biologist who will start his first lab at Kansas State University in Manhattan this month. Parthasarathy says that reviewers should take a step back and think about what is reasonable to expect as proof in a single paper.

CONSISTENT COMMUNICATION

When a revision comes back, reviewers must be consistent. For example, if the authors added an experiment, it would be unreasonable for reviewers to ask for yet more data. "Authors report that they feel that the goalposts are moved during revisions, which gets

frustrating,” says Wendy Lipworth, a sociologist who has studied the ethics of scientific review at the University of Sydney in Australia.

The most frustrating rookie offence, however, might be making contradictory assessments in a single review. Kulp says it drives editors “insane” when a reviewer submits highly critical comments with a recommendation to “publish as is”. Such reviews are most common when journals allow reviewers to submit one set of comments to the editor and another to authors. Contrasting reviews create problems for everybody concerned, says Parthasarathy. At best, they make the editor’s decision harder; at worst, the catty ones can start a feud. Reviewers should never write anything that would be damaging if their identity were revealed.

Reviewers should avoid overestimating their own capacity to review multidisciplinary papers: a lack of understanding could lead the reviewer to recommend that perfectly good work is rejected, says Erik De Schutter, a theoretical neurobiologist at the Okinawa Institute of Science and Technology in Japan. He is editor-in-chief for *Neuroinformatics*, where he has published a paper on his difficulties in getting theoretical modelling papers accepted in general neuroscience journals, many of which insisted on experimental data (E. De Schutter, *Neuroinformatics* 6, 253–255; 2008). Being qualified to make comments on only one part of a paper doesn’t rule out valuable contributions, but reviewers should be open about where their expertise lies. “I am a theorist, so quite often when I’m asked to review experimental papers, I make clear that I’m not qualified to judge the methods,” says De Schutter.

Whatever the content of the review, referees should be completely honest about their affiliations and who helped them to write it. Often, senior scientists invite graduate students or postdocs to write or contribute to the review. Most editors don’t mind this, as long as they are notified in advance and all contributors are listed. “Finding the proper mix of expertise among reviewers is a careful calibration on the part of the manuscript editor — one that can get screwed up if the reviewer is not who you think it is,” says Parthasarathy. Graduate students have every right to make sure that their mentors explain how they have helped.

Conflicts of interest may sometimes dictate that researchers decline a review — for

example, if they have financial stakes in the paper’s content or personal ties with the author. Some journals don’t mind, as long as all issues are disclosed. “Biases are not all conflicts of interest; and sometimes editors want the perspective of someone inclined to loathe a particular piece of work,” says Lipworth. If, for example, a potential reviewer is actively writing a paper on the same topic, says Sage, it is probably best to decline the review. But if the reviewer’s own paper has already been submitted, and therefore is documented in the scientific record, the reviewer can’t be accused of stealing ideas.

GETTING NOTICED

The most straightforward way for researchers to become peer reviewers is for their mentors to introduce them to editors, but there are other routes. The best is to do good science and get published; manuscript editors scour citation databases and conferences, looking for young scientists with expertise that might make them a valued reviewer. Scientific meetings are the most appropriate place for would-be reviewers to introduce themselves to editors. In fact, finding bright young reviewers is one of the main reasons journal editors attend meetings. E-mailing an editor out of the blue is a riskier endeavour and can be seen as overly pushy, says Sage.

Novice reviewers might also find that journals run by scientific societies are often short on reviewers, and so are eager to get them involved. The Ecological Society of America in Washington DC, publisher of *Ecological Applications*, contacts its student members to find reviewers. Others, such as the *American Journal of Pharmaceutical Education*, encourage potential reviewers to sign up on their website. “I don’t have many reviewers for the pharmacokinetics papers so I’ve been pushing to get people engaged,” says Gayle Brazeau, the journal’s associate editor.

Reviewing may seem like a time sink in the short term, but it can have long-term benefits. For one, watching manuscripts evolve through the editorial process can be a valuable experience, says Veeman. And editors sometimes reward thoughtful, articulate reviews with further opportunities. Schimel says that excellent reviewers might be offered seats on a journal’s editorial advisory board. And, notes Wacek, reviewing can be an entry point to a career in scientific publishing.

Spigt’s reviews helped him to get a position as an assistant editor with the *Journal of Clinical Epidemiology*, a leading publication in his field. He hopes one day to become an associate editor. “One of the most important things manuscript reviews can do,” says Spigt, “is help the reviewer build relationships with the journals in which they want to publish.” ■

Virginia Gewin is a journalist based in Portland, Oregon.



“Reviewing manuscripts makes me feel like I’m a fully fledged member of the scientific community.”

Claudio Casola

UNITED STATES

Postdocs probe industry

A programme launched this year at the University of California, Berkeley, aims to give its postdocs an inside view of local life-sciences companies and other businesses. The Postdoc Industry Exploration Program (PIEP) was instituted after Berkeley postdocs ran a successful pilot programme. The PIEP will be offered every year to Berkeley’s 1,100 science postdocs, and possibly also to its graduate students. PIEP participants meet company researchers, administrators and executives, learn about their work and establish alliances that could lead to job offers. More than 100 postdocs joined the pilot: half indicated an interest in industry before the programme began, whereas three-quarters did after it ended. Christopher Tsang, PIEP co-founder and a postdoc at Berkeley, says that the project will be shared with postdocs at other institutions at next year’s US National Postdoc Association meeting.

EUROPE

PhD funding inadequate

Funding for doctoral candidates in some European nations often runs out before the research projects end and doesn’t cover living expenses, according to a 2008–09 survey of more than 7,500 PhD students from 12 countries. In Spain, 31% of respondents said that funding does not adequately cover their living costs; in Portugal it was 24%, and in Croatia 23%. The report, by the European Council of Doctoral Candidates and Junior Researchers (Eurodoc) in Brussels, was released on 30 September. Eurodoc’s most extensive survey to date, the report documents trends in career paths, funding, mobility, training and working conditions.

UNITED KINGDOM

Science careers unstable

UK scientists are concerned about career instability and lack of research positions, says a poll by an advocacy group, Science is Vital. It surveyed 700 science PhD students and researchers in September at the behest of UK science minister David Willetts, who met group leaders on 6 October and is arranging a discussion with government and funding representatives. Respondents cited problems with short-term contracts, low pay, compromised mobility and lack of work–life balance. Many want smaller labs and permanent academic posts funded by universities.

frustrating,” says Wendy Lipworth, a sociologist who has studied the ethics of scientific review at the University of Sydney in Australia.

The most frustrating rookie offence, however, might be making contradictory assessments in a single review. Kulp says it drives editors “insane” when a reviewer submits highly critical comments with a recommendation to “publish as is”. Such reviews are most common when journals allow reviewers to submit one set of comments to the editor and another to authors. Contrasting reviews create problems for everybody concerned, says Parthasarathy. At best, they make the editor’s decision harder; at worst, the catty ones can start a feud. Reviewers should never write anything that would be damaging if their identity were revealed.

Reviewers should avoid overestimating their own capacity to review multidisciplinary papers: a lack of understanding could lead the reviewer to recommend that perfectly good work is rejected, says Erik De Schutter, a theoretical neurobiologist at the Okinawa Institute of Science and Technology in Japan. He is editor-in-chief for *Neuroinformatics*, where he has published a paper on his difficulties in getting theoretical modelling papers accepted in general neuroscience journals, many of which insisted on experimental data (E. De Schutter, *Neuroinformatics* 6, 253–255; 2008). Being qualified to make comments on only one part of a paper doesn’t rule out valuable contributions, but reviewers should be open about where their expertise lies. “I am a theorist, so quite often when I’m asked to review experimental papers, I make clear that I’m not qualified to judge the methods,” says De Schutter.

Whatever the content of the review, referees should be completely honest about their affiliations and who helped them to write it. Often, senior scientists invite graduate students or postdocs to write or contribute to the review. Most editors don’t mind this, as long as they are notified in advance and all contributors are listed. “Finding the proper mix of expertise among reviewers is a careful calibration on the part of the manuscript editor — one that can get screwed up if the reviewer is not who you think it is,” says Parthasarathy. Graduate students have every right to make sure that their mentors explain how they have helped.

Conflicts of interest may sometimes dictate that researchers decline a review — for

example, if they have financial stakes in the paper’s content or personal ties with the author. Some journals don’t mind, as long as all issues are disclosed. “Biases are not all conflicts of interest; and sometimes editors want the perspective of someone inclined to loathe a particular piece of work,” says Lipworth. If, for example, a potential reviewer is actively writing a paper on the same topic, says Sage, it is probably best to decline the review. But if the reviewer’s own paper has already been submitted, and therefore is documented in the scientific record, the reviewer can’t be accused of stealing ideas.

GETTING NOTICED

The most straightforward way for researchers to become peer reviewers is for their mentors to introduce them to editors, but there are other routes. The best is to do good science and get published; manuscript editors scour citation databases and conferences, looking for young scientists with expertise that might make them a valued reviewer. Scientific meetings are the most appropriate place for would-be reviewers to introduce themselves to editors. In fact, finding bright young reviewers is one of the main reasons journal editors attend meetings. E-mailing an editor out of the blue is a riskier endeavour and can be seen as overly pushy, says Sage.

Novice reviewers might also find that journals run by scientific societies are often short on reviewers, and so are eager to get them involved. The Ecological Society of America in Washington DC, publisher of *Ecological Applications*, contacts its student members to find reviewers. Others, such as the *American Journal of Pharmaceutical Education*, encourage potential reviewers to sign up on their website. “I don’t have many reviewers for the pharmacokinetics papers so I’ve been pushing to get people engaged,” says Gayle Brazeau, the journal’s associate editor.

Reviewing may seem like a time sink in the short term, but it can have long-term benefits. For one, watching manuscripts evolve through the editorial process can be a valuable experience, says Veeman. And editors sometimes reward thoughtful, articulate reviews with further opportunities. Schimel says that excellent reviewers might be offered seats on a journal’s editorial advisory board. And, notes Wacek, reviewing can be an entry point to a career in scientific publishing.

Spigt’s reviews helped him to get a position as an assistant editor with the *Journal of Clinical Epidemiology*, a leading publication in his field. He hopes one day to become an associate editor. “One of the most important things manuscript reviews can do,” says Spigt, “is help the reviewer build relationships with the journals in which they want to publish.” ■

Virginia Gewin is a journalist based in Portland, Oregon.



“Reviewing manuscripts makes me feel like I’m a fully fledged member of the scientific community.”

Claudio Casola

UNITED STATES

Postdocs probe industry

A programme launched this year at the University of California, Berkeley, aims to give its postdocs an inside view of local life-sciences companies and other businesses. The Postdoc Industry Exploration Program (PIEP) was instituted after Berkeley postdocs ran a successful pilot programme. The PIEP will be offered every year to Berkeley’s 1,100 science postdocs, and possibly also to its graduate students. PIEP participants meet company researchers, administrators and executives, learn about their work and establish alliances that could lead to job offers. More than 100 postdocs joined the pilot: half indicated an interest in industry before the programme began, whereas three-quarters did after it ended. Christopher Tsang, PIEP co-founder and a postdoc at Berkeley, says that the project will be shared with postdocs at other institutions at next year’s US National Postdoc Association meeting.

EUROPE

PhD funding inadequate

Funding for doctoral candidates in some European nations often runs out before the research projects end and doesn’t cover living expenses, according to a 2008–09 survey of more than 7,500 PhD students from 12 countries. In Spain, 31% of respondents said that funding does not adequately cover their living costs; in Portugal it was 24%, and in Croatia 23%. The report, by the European Council of Doctoral Candidates and Junior Researchers (Eurodoc) in Brussels, was released on 30 September. Eurodoc’s most extensive survey to date, the report documents trends in career paths, funding, mobility, training and working conditions.

UNITED KINGDOM

Science careers unstable

UK scientists are concerned about career instability and lack of research positions, says a poll by an advocacy group, Science is Vital. It surveyed 700 science PhD students and researchers in September at the behest of UK science minister David Willetts, who met group leaders on 6 October and is arranging a discussion with government and funding representatives. Respondents cited problems with short-term contracts, low pay, compromised mobility and lack of work–life balance. Many want smaller labs and permanent academic posts funded by universities.

COMPLAINTS DEPARTMENT

Reap what you sow.

BY THORAIYA DYER

The queue was long again, but Elise couldn't stay at home watching her crop shrink and shrivel while she waited for an electronic reply from the QuickLife complaints department.

Mud and straw were caked onto her boots. She clicked her heels surreptitiously. Some of the mud flaked off. The man in front of her smelled of poultry. He slowly scratched the sunburn on the back of his neck.

At last, she was led into an interview room where a younger woman with a headset and fuchsia lips waved her to a chair.

"You asked to see a supervisor, Miss Green. I am Tanaivin Biss."

"And I am angry," Elise said, the wait not having improved her temper, "at the way QuickLife is attempting to acquire my land."

"Please state your complaint as either a product fault or a customer service delay."

Elise leaned back in her chair. She plonked her boots, one at a time, onto Tanaivin Biss's glass desktop and waited for the shower of mud flakes to settle.

"Let's call it a product fault. The seeds that you sold me leached all the phosphorus from my soil."

They'd set it all up for her. The latest in biotech: accelerated wheat. They told her it was within her budget, that using accelerated wheat would lead to a tenfold increase in production. Nine days from germination to ripe kernels at maximum dry weight. She'd already made two complete harvests, investing the cash in a new truck for more efficient transport.

But this time, there was nothing to harvest. When the sun had come up on the morning of the ninth day, the straggly, mostly immature plants were intermittently laced with the telltale purple tint of phosphorus deficiency. The few heads that had formed were small, the grains shrivelled. Feed quality only.

Biss subvocalized and was answered by her headset.

"Our records indicate your soil tested high for phosphorus."

"Are you telling me QuickLife can't interpret its own results? It may be high but it sure isn't available."

"Wait. Yes. Unfortunately, the full details of your new fertilizer schedule were not provided at the time of your purchase. The computer is telling me that an application of superphosphate between crops would be advantageous."



"Every ten days? I'll tell you what would be advantageous. A free crop-dusting plane and pilot!"

The door opened behind her, but Elise didn't turn around.

"Will there be anything else, Miss Green?" Biss said.

"Do you have any idea how much aerial topdressing with superphosphate costs? I can't afford it unless you increase the limit of my loan."

"Your account with QuickLife's lending branch is already overdrawn."

"I can't believe this. It wasn't even my idea!"

"When you visited the complaints department three weeks ago, you said you were expecting legal action following the germination of a QuickLife product in a neighbouring field."

"Yes. My wheat is GM. Theirs is organic."

"Increased production is only a side effect of acceleration. The main feature of an accelerated strain is that the flowering and fruiting phase occur within such a narrow window that contamination of neighbouring, non-GM crops is negligible to nil."

"My harvest will be negligible to nil if you don't refund my money or extend my overdraft so I can pay for fertilizer."

"I'm sorry, Miss Green. I am unable to assist you."

Elise stared at the vending machine with its bottled spring-water for a full minute before marching outside, detaching the garden bed watering system from the rainwater tank, and slaking her thirst at the tap.

A moment later, she became aware of somebody sitting despondently on the edge of the concrete turning circle. It was the man

who had gone in ahead of her.

"No luck?" she ventured.

"None at all."

She liked the deep, rich roll of his voice. Wiping her wet palms on her overalls, she sat next to him.

"What happened?"

"Fifty thousand accelerated geese ate me out of house and home. Twenty-four hours, they give me. Just one day to remove the bloody birds from my farm what QuickLife just took."

Elise's heart thumped excitedly against her ribs.

"I have fifty thousand acres," she said. "How do geese go on phosphorus-deficient wheat grass?"

"All my geese got micronutrient implants."

Elise stuck out her hand.

"I'm Elise Green."

He shook solemnly. His skin was warm and dry like bread from the oven.

"Peter Cross."

"My house," she said, "has a spare room. And I have a new truck. This way."

He moved into her house, and the haste with which their hearts became entangled might have been unseemly if they hadn't been surrounded by accelerated geese feasting on accelerated wheat and leaving phosphate-rich deposits all over Elise's 50,000 acres at a frighteningly accelerated rate.

Elise soon discovered that although her imagined, strong-and-silent-type ideal was embodied by Peter Cross, cohabitation turned out to be only marginally less lonely than living alone. By the time Peter's geese had been sent to market, allowing him to purchase a small townhouse and take up work at the accelerated goose hatchery, they had already parted ways amicably.

She called him, sometimes. She liked the sound of his voice.

"They have to feed these little buggers while they're still in the egg," Peter marvelled. "They suck the yolks out of half a dozen normal eggs just to get one accelerated egg to hatch."

"It's a little bit boring," Elise confessed, "waiting for normal wheat to ripen."

And she tinkered with her irrigators as she watched the sun go down behind the whispering hills. ■

Thoraiya Dyer, winner of the 2011 Ditmar Award for Best New Talent, is an Australian veterinarian and speculative fiction writer with a healthy respect for hungry geese. www.thoraiyadyer.com.

➔ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

InsP₃R channel gating altered by clustering?

ARISING FROM T. Rahman, A. Skupin, M. Falcke & C. W. Taylor *Nature* **458**, 655–659 (2009)

The inositol trisphosphate receptor (InsP₃R) forms a calcium channel that resides in the membrane of the endoplasmic reticulum and is activated by inositol trisphosphate (InsP₃). InsP₃ is a phosphorylated monosaccharide that is generated via hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP₂), a phospholipid that is located in the plasma membrane, and activation of the InsP₃R is involved in a broad range of biological processes, including cell division, apoptosis and development. Rahman *et al.*^{1,2} reported that exposure to low concentrations of InsP₃ induces rapid clustering of InsP₃R Ca²⁺ release channels normally randomly distributed in endoplasmic reticulum/outer nuclear membranes. Importantly, clustered channels gate differently from lone channels. Using similar protocols, we observed InsP₃R channel clustering without exposure to InsP₃ (Fig. 1a), as we found in other systems^{3–5} with protocols designed to avoid InsP₃ pre-exposure. More significantly, we find that clustering has no effect on InsP₃R channel gating. For this reason, we believe that InsP₃-induced channel clustering and modification of channel gating by clustering may not be universal phenomena.

Rahman *et al.*^{1,2} reported that in sub-optimal cytoplasmic free Ca²⁺ concentrations ([Ca²⁺]_i), clustered recombinant rat type 3 InsP₃R (InsP₃R3) channels expressed in InsP₃R-deficient DT40-KO cells gated identically and independently, but with lower open probability

(*P*_o) than lone channels, regardless of cluster size. In contrast, clustered channels had the same *P*_o as lone channels in optimal ligand conditions, but gated with positive cooperativity. If broadly observed, these surprising findings have important implications for understanding InsP₃-mediated Ca²⁺ signals, and for quantitative analyses in single-channel InsP₃R electrophysiology.

To verify these observations, we examined the same InsP₃R3 channels in the same DT40-KO cells using similar protocols and ligand conditions. Specifically, we used 5 mM (same as Rahman *et al.*¹) and 0.5 mM (more physiological) cytoplasmic free [ATP⁴⁻] ([ATP]_i). Records with ≤4 active channels were analysed with the same algorithm¹. In addition, we similarly analysed nuclear patch-clamp records previously acquired under comparable ligand conditions for recombinant rat InsP₃R3 expressed in *Xenopus* oocytes⁶, endogenous *Xenopus* type 1 InsP₃R (InsP₃R1) in oocytes⁷ and endogenous insect InsP₃R in Sf9 cells⁵. For all channels examined in these various systems, we detected no statistical difference (*P* > 0.05, *t*-test) between *P*_o in single- versus multi-channel patches in saturating [InsP₃] and sub-optimal [Ca²⁺]_i (Fig. 1b), or in sub-saturating [InsP₃] and optimal [Ca²⁺]_i (Fig. 1c). Furthermore, in two-channel records, similar channel gating patterns were detected in all [Ca²⁺]_i (Fig. 2), with only a small fraction exhibiting positive cooperativity. Thus, our extensive data set reveals no effect of clustering on InsP₃R channel gating in all ligand conditions.

In constant ligand conditions, we consistently observed abrupt, stochastic, irreversible inactivation of InsP₃R in on-nucleus or excised luminal-side-out nuclear patches, with mean activity durations of ~40 s for oocyte InsP₃R (ref. 3), ~100 s for Sf9 InsP₃R (ref. 5) and ~140 s for InsP₃R from DT40-KO cells, whereas Rahman *et al.* reported no such inactivation¹. Importantly, we analysed only current records long enough for the number of active channels to be counted with >99% confidence^{1,5,6,8}. Because finite time elapsed between pipettes making contact with the outer nuclear membrane and gigaohm seal formation (<5 s for oocyte and Sf9 nuclei, ~10 s for DT40 nuclei), apparent single-channel patches possibly included a fraction (~11–26%) that actually contained multiple channels in

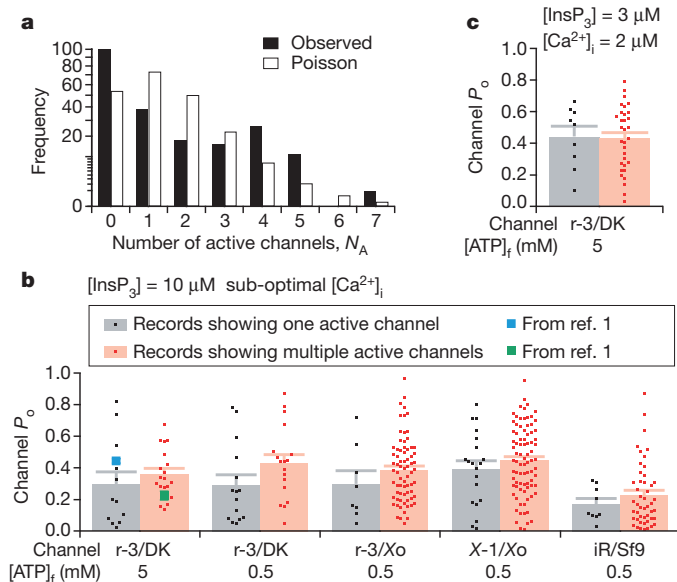


Figure 1 | InsP₃R channels are clustered before exposure to InsP₃, with gating properties unaltered by clustering. **a**, N_A in nuclear membrane patches with no pre-exposure to InsP₃ obtained from InsP₃R3 expressing DT40-KO cells. Note nonlinear square-root scale for frequency axis. **b**, P_o observed under saturating [InsP₃] and sub-optimal [Ca²⁺]_i in multi- and apparent single-channel current records for recombinant rat InsP₃R3 (r-3) channels expressed in DT40-KO (DK) cells or *Xenopus* oocytes (Xo), endogenous *Xenopus* InsP₃R1 (X-1) channels from *Xenopus* oocytes (Xo), and endogenous insect InsP₃R (iR) channels from Sf9 cells. Concentrations of free ATP⁴⁻ ([ATP]_i) in the pipette solutions used are indicated. Mean P_o with s.e.m. (as error bars) and P_o for individual current records are shown, together with mean P_o from Rahman *et al.*¹ **c**, P_o of r-3 channels in DK cells in optimal [Ca²⁺]_i and sub-saturating [InsP₃], ligand conditions not investigated in Rahman *et al.*¹. Same symbols as in **b** are used.

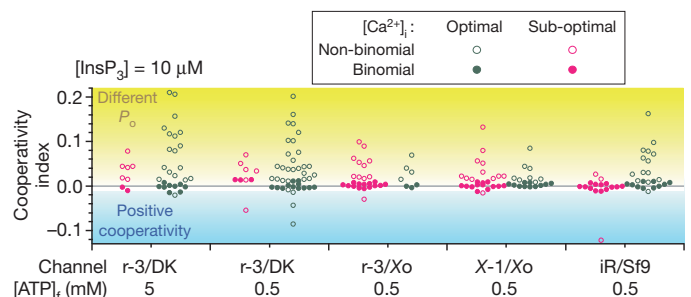


Figure 2 | Distribution of cooperativity index for two-channel current records of different InsP₃R channels in various systems in optimal and sub-optimal [Ca²⁺]_i. Filled and open circles represent records with two channels exhibiting identical and independent, or non-binomial gating, respectively. Non-binomial records with cooperativity index, $(P_2 + P_1/2) - P_2$, significantly greater than 0 (in yellow shaded region) had two channels gating with different P_o , and those with cooperativity index significantly smaller than 0 (in blue shaded region) had two channels gating with positive cooperativity. The cooperativity indices have no correlation with the durations of the current records (data not shown) and therefore are unlikely to be significantly affected by current record durations limited by channel inactivation.

which all but one channel inactivated before gigaohm seal formation. Yet, the mean number of active channels (N_A) we observed for InsP₃R3 in DT40 nuclear patches (1.36 ± 0.12 for 211 patches) is similar to that reported in Rahman *et al.*¹, suggesting that inactivation did not substantially impair our ability to count channels in these patches. We did detect larger N_A (10.8 ± 1)⁸ in outside-out nuclear patches, which probably have significantly larger membrane areas and were isolated using a different technique, and therefore are not an appropriate comparison to illustrate either the variability in InsP₃R expression level in DT40-KO cells or the effect of inactivation on N_A detected.

The P_o distributions that we observed in apparent single-channel and true multi-channel patches were similar, with no indication that two populations of channels with different P_o exist. Furthermore, mean P_o of our true multi-channel patches is comparable to the lone-channel P_o observed by Rahman *et al.*¹ (Fig. 1b). Thus, our conclusion that clustering does not affect InsP₃R channel gating is not compromised by the irreversible inactivation of InsP₃R channels.

We have no clear explanation for the discrepancies between our observations and those reported by Rahman *et al.*¹. However, neither InsP₃-induced InsP₃R clustering nor its modification of InsP₃R gating are consistently observed for InsP₃R expressed in DT40-KO and other cells. In contrast, channel clustering before InsP₃ exposure was observed in all cell systems investigated without effect on channel gating^{3–5}. Thus, we suggest that InsP₃-induced channel clustering and modification of channel gating by clustering may not be universal phenomena.

METHODS SUMMARY

Single-channel P_o for a sufficiently long⁵ current record with N_A active channels was evaluated as $= \sum_{i=0}^{N_A} [iP_i] / N_A$, where P_i , the probability that i channels were active simultaneously in the record, was determined by the same method as Rahman *et al.*¹. The channel gating pattern for a two-channel current record was determined from P_i . If P_i are similar to the expected binomial values, $[2! P_o^i (1 - P_o)^{2-i}] / [i! (2-i)!]$ for $i=0, 1, 2$ ($P > 0.05$ by χ^2 -test), the channels gated independently with similar P_o . Otherwise, if the cooperativity index,

$(P_2 + P_1/2)^2 - P_2$, is >0 , they gated with different P_o , with or without negative cooperativity. If $(P_2 + P_1/2)^2 - P_2 < 0$, they gated with positive cooperativity⁹.

Horia Vais¹, J. Kevin Foskett^{1,2} & Don-On Daniel Mak¹

¹Department of Physiology, University of Pennsylvania Perleman School of Medicine, Philadelphia, Pennsylvania 19104, USA.

e-mail: dmak@mail.med.upenn.edu

²Department of Cell and Developmental Biology, University of Pennsylvania Perleman School of Medicine, Philadelphia, Pennsylvania 19104, USA.

Received 20 October 2010; accepted 11 August 2011.

1. Rahman, T. U., Skupin, A., Falcke, M. & Taylor, C. W. Clustering of InsP₃ receptors by InsP₃ retunes their regulation by InsP₃ and Ca²⁺. *Nature* **458**, 655–659 (2009).
2. Rahman, T. & Taylor, C. W. Dynamic regulation of IP₃ receptor clustering and activity by IP₃. *Channels (Austin)* **3**, 226–232 (2009).
3. Mak, D.-O. D. & Foskett, J. K. Single-channel kinetics, inactivation, and spatial distribution of inositol trisphosphate (IP₃) receptors in *Xenopus* oocyte nucleus. *J. Gen. Physiol.* **109**, 571–587 (1997).
4. Mak, D.-O. D. *et al.* Single-channel properties in endoplasmic reticulum membrane of recombinant type 3 inositol trisphosphate receptor. *J. Gen. Physiol.* **115**, 241–256 (2000).
5. Ionescu, L. *et al.* Graded recruitment and inactivation of single InsP₃ receptor Ca²⁺-release channels: implications for quantal Ca²⁺ release. *J. Physiol. (Lond.)* **573**, 645–662 (2006).
6. Mak, D.-O. D., McBride, S. & Foskett, J. K. Regulation by Ca²⁺ and inositol 1,4,5-trisphosphate (InsP₃) of single recombinant type 3 InsP₃ receptor channels. Ca²⁺ activation uniquely distinguishes types 1 and 3 InsP₃ receptors. *J. Gen. Physiol.* **117**, 435–446 (2001).
7. Mak, D.-O. D., McBride, S. & Foskett, J. K. Inositol 1,4,5-trisphosphate activation of inositol trisphosphate receptor Ca²⁺ channel by ligand tuning of Ca²⁺ inhibition. *Proc. Natl Acad. Sci. USA* **95**, 15821–15825 (1998).
8. Vais, H. *et al.* Redox-regulated heterogeneous thresholds for ligand recruitment among InsP₃R Ca²⁺ release channels. *Biophys. J.* **99**, 407–416 (2010).
9. Kenyon, J. L. & Bauer, R. J. Amplitude histograms can identify positively but not negatively coupled channels. *J. Neurosci. Methods* **96**, 105–111 (2000).

Author Contributions All authors contributed to project planning and wrote the paper. H.V. and D.-O.D.M. did data acquisition and analysis.

Competing financial interests: declared none.

doi:10.1038/nature10493

Rahman *et al.* reply

REPLYING TO H. Vais, J. K. Foskett & D.-O. D. Mak *Nature* **478**, doi:10.1038/nature10493 (2011)

We observe that under conditions designed to mimic those of an unstimulated cell, InsP₃ causes InsP₃R to assemble rapidly and reversibly into small clusters of about 4–5 channels, and that gating of lone and clustered InsP₃R by InsP₃ and Ca²⁺ is different. We speculate that dynamic assembly of InsP₃R clusters by InsP₃ may contribute to the genesis of elementary Ca²⁺ release events by both assembling clusters and re-tuning the behaviour of InsP₃R within them^{1–3}. Vais *et al.*⁴ report that they have been unable to replicate some of our findings.

We cannot explain their results, but we note the considerably greater variability in their observations (Fig. 1a). This may be related to difficulties in reliably estimating the number of InsP₃R (N_A) within a patch. They reported that under the conditions used in their figure 1 (ref. 4), “abrupt, irreversible channel inactivation was consistently observed”⁵ in Sf9 cells^{5,6}, DT40 cells^{5,7} and *Xenopus* oocytes⁷, with time constants of ~30–120 s (ref. 7). We observe no inactivation¹. This loss of activity, before and after seal formation⁶, may cause unreliable estimation of N_A . Indeed, in patches isolated from the same DT40 nuclei, they report N_A of 1.36 for lumen-out patches⁴ but 10.8 for cytosol-out patches where inactivation is much slower⁵. Results in their

figure 1b (ref. 4) show some patches in which open probability (P_o) is <0.02 . Recordings (>65 s) comparable to the time course of InsP₃R inactivation⁴ are required to establish confidently that such patches contain only a single channel^{1,6}.

Vais *et al.* also suggest elsewhere that N_A increases with the intensity of InsP₃R stimulation^{5,6}, and that the effect is modulated by sulphhydryl reagents⁵. Our limited analysis failed to replicate this³, but it is noteworthy that our analyses¹ and theirs⁴ differ in the amounts of 2-mercaptoethanol included in culture media. Considerable variation in the amplitude of InsP₃R3 unitary currents^{5,8} and the frequent occurrence of sub-conducting states within records⁸ may also confound their estimates of N_A . Counting InsP₃R (figure 1a in ref. 4), calculating P_o (figure 1b, c in ref. 4), and assessing whether P_o is binomially distributed (figure 2 in ref. 4) depend on reliably determining N_A . Our recordings are stable for many minutes with no inactivation or InsP₃R recruitment, and only extremely rare visits to sub-conducting states^{1,2}. We suggest that the more complex behaviour of InsP₃R observed by Vais *et al.*⁴ may compromise estimates of N_A , so that even patches in which they detect only one active InsP₃R may

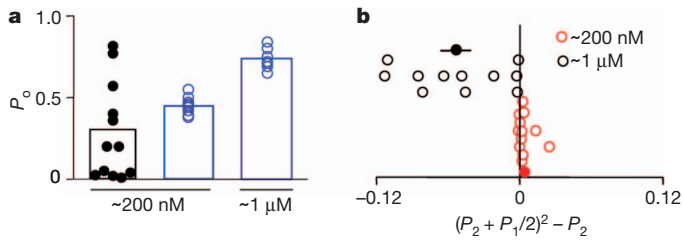


Figure 1 | InsP₃-evoked clustering of InsP₃R alters their gating. **a**, InsP₃R3 expressed in DT40-KO cells were stimulated with 10 μ M InsP₃ in cytosol-like medium containing a free $[Ca^{2+}]$ of ~ 200 nM or ~ 1 μ M. P_o is shown for patches with a single InsP₃R (blue) taken from our results¹ and for equivalent analyses taken from figure 1b of Vais *et al.*⁴ (black). **b**, The analysis reported in ref. 4 was used to calculate the probabilities of 1 (P_1) or 2 (P_2) InsP₃R being open in patches with 2 InsP₃R after stimulation with 10 μ M InsP₃ and the indicated concentrations of free $[Ca^{2+}]$. We applied their analysis⁴ to determine whether InsP₃R gated independently or cooperatively by computing $(P_2 + P_1/2)^2 - P_2$, where $(P_2 + P_1/2)$ provides the estimate of P_o . Results are shown for each of 12 and 11 patches with 200 nM and 1 μ M free $[Ca^{2+}]$, respectively (open symbols); and as mean \pm s.e.m. (solid symbols: 0.0048 ± 0.0011 and -0.0517 ± 0.0127 , for 200 nM and 1 μ M $[Ca^{2+}]$, respectively; error bars are smaller than the symbol for the former). The results are consistent with independent gating of InsP₃R in 200 nM free $[Ca^{2+}]$, and with positively cooperative gating in 1 μ M free $[Ca^{2+}]$.

contain several inactivated InsP₃R that may cluster and regulate a neighbour¹.

Finally, and to allow direct comparison with the analysis Vais *et al.* show in figure 2 of ref. 4, we re-analysed our data¹ and included results from InsP₃R1 (see Fig. 1b). This provides statistical evidence that in patches with two InsP₃R, InsP₃R gate independently at resting $[Ca^{2+}]$ but with positive cooperativity at elevated $[Ca^{2+}]$.

Taufiq Rahman¹, Alexander Skupin^{2,3}, Martin Falcke⁴ & Colin W. Taylor¹

¹Department of Pharmacology, Tennis Court Road, Cambridge CB2 1PD, UK.

e-mail: cwt1000@cam.ac.uk

²Luxembourg Centre for Systems Biomedicine, 162a Avenue de la Faïencerie, University of Luxembourg, L-1511, Luxembourg.

³Institute for Systems Biology, 401 Terry Avenue North, Seattle, Washington 98103, USA.

⁴Mathematical Cell Physiology, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str, 10, 13092 Berlin, Germany.

1. Rahman, T. U., Skupin, A., Falcke, M. & Taylor, C. W. Clustering of InsP₃ receptors by InsP₃ retunes their regulation by InsP₃ and Ca^{2+} . *Nature* **458**, 655–659 (2009).
2. Rahman, T. & Taylor, C. W. Dynamic regulation of IP₃ receptor clustering and activity by IP₃. *Channels* **3**, 226–232 (2009).
3. Taylor, C. W. & Tovey, S. C. IP₃ receptors: toward understanding their activation. *Cold Spring Harb. Perspect. Biol.* **2**, a004010 (2010).
4. Vais, H., Foskett, J. K. & Mak, D.-O. D. InsP₃R channel gating altered by clustering? *Nature* **478**, doi:10.1038/nature10493 (2011).
5. Vais, H. *et al.* Redox-regulated heterogeneous thresholds for ligand recruitment among InsP₃R Ca^{2+} -release channels. *Biophys. J.* **99**, 407–416 (2010).
6. Ionescu, L. *et al.* Graded recruitment and inactivation of single InsP₃ receptor Ca^{2+} -release channels: implications for quantal Ca^{2+} release. *J. Physiol. (Lond.)* **573**, 645–662 (2006).
7. Mak, D.-O. D., White, C., Ionescu, L. & Foskett, J. K. in *Calcium Signaling* (ed. Putney, J. W. Jr) 203–229 (Taylor and Francis, 2006).
8. Vais, H., Foskett, J. K. & Mak, D. O. Unitary Ca^{2+} current through recombinant type 3 InsP₃ receptor channels under physiological ionic conditions. *J. Gen. Physiol.* **136**, 687–700 (2010).

Author Contributions T.R. performed the experiments and, with C.W.T., analysed the data. C.W.T. and T.R. wrote the paper with input from A.S. and M.F. The project was directed by C.W.T.

doi:10.1038/nature10494

InsP₃R channel gating altered by clustering?

ARISING FROM T. Rahman, A. Skupin, M. Falcke & C. W. Taylor *Nature* **458**, 655–659 (2009)

The inositol trisphosphate receptor (InsP₃R) forms a calcium channel that resides in the membrane of the endoplasmic reticulum and is activated by inositol trisphosphate (InsP₃). InsP₃ is a phosphorylated monosaccharide that is generated via hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP₂), a phospholipid that is located in the plasma membrane, and activation of the InsP₃R is involved in a broad range of biological processes, including cell division, apoptosis and development. Rahman *et al.*^{1,2} reported that exposure to low concentrations of InsP₃ induces rapid clustering of InsP₃R Ca²⁺ release channels normally randomly distributed in endoplasmic reticulum/outer nuclear membranes. Importantly, clustered channels gate differently from lone channels. Using similar protocols, we observed InsP₃R channel clustering without exposure to InsP₃ (Fig. 1a), as we found in other systems^{3–5} with protocols designed to avoid InsP₃ pre-exposure. More significantly, we find that clustering has no effect on InsP₃R channel gating. For this reason, we believe that InsP₃-induced channel clustering and modification of channel gating by clustering may not be universal phenomena.

Rahman *et al.*^{1,2} reported that in sub-optimal cytoplasmic free Ca²⁺ concentrations ([Ca²⁺]_i), clustered recombinant rat type 3 InsP₃R (InsP₃R3) channels expressed in InsP₃R-deficient DT40-KO cells gated identically and independently, but with lower open probability

(P_o) than lone channels, regardless of cluster size. In contrast, clustered channels had the same P_o as lone channels in optimal ligand conditions, but gated with positive cooperativity. If broadly observed, these surprising findings have important implications for understanding InsP₃-mediated Ca²⁺ signals, and for quantitative analyses in single-channel InsP₃R electrophysiology.

To verify these observations, we examined the same InsP₃R3 channels in the same DT40-KO cells using similar protocols and ligand conditions. Specifically, we used 5 mM (same as Rahman *et al.*¹) and 0.5 mM (more physiological) cytoplasmic free [ATP⁴⁻] ([ATP]_i). Records with ≤4 active channels were analysed with the same algorithm¹. In addition, we similarly analysed nuclear patch-clamp records previously acquired under comparable ligand conditions for recombinant rat InsP₃R3 expressed in *Xenopus* oocytes⁶, endogenous *Xenopus* type 1 InsP₃R (InsP₃R1) in oocytes⁷ and endogenous insect InsP₃R in Sf9 cells⁵. For all channels examined in these various systems, we detected no statistical difference (P > 0.05, *t*-test) between P_o in single- versus multi-channel patches in saturating [InsP₃] and sub-optimal [Ca²⁺]_i (Fig. 1b), or in sub-saturating [InsP₃] and optimal [Ca²⁺]_i (Fig. 1c). Furthermore, in two-channel records, similar channel gating patterns were detected in all [Ca²⁺]_i (Fig. 2), with only a small fraction exhibiting positive cooperativity. Thus, our extensive data set reveals no effect of clustering on InsP₃R channel gating in all ligand conditions.

In constant ligand conditions, we consistently observed abrupt, stochastic, irreversible inactivation of InsP₃R in on-nucleus or excised luminal-side-out nuclear patches, with mean activity durations of ~40 s for oocyte InsP₃R (ref. 3), ~100 s for Sf9 InsP₃R (ref. 5) and ~140 s for InsP₃R from DT40-KO cells, whereas Rahman *et al.* reported no such inactivation¹. Importantly, we analysed only current records long enough for the number of active channels to be counted with >99% confidence^{1,5,6,8}. Because finite time elapsed between pipettes making contact with the outer nuclear membrane and gigaohm seal formation (<5 s for oocyte and Sf9 nuclei, ~10 s for DT40 nuclei), apparent single-channel patches possibly included a fraction (~11–26%) that actually contained multiple channels in

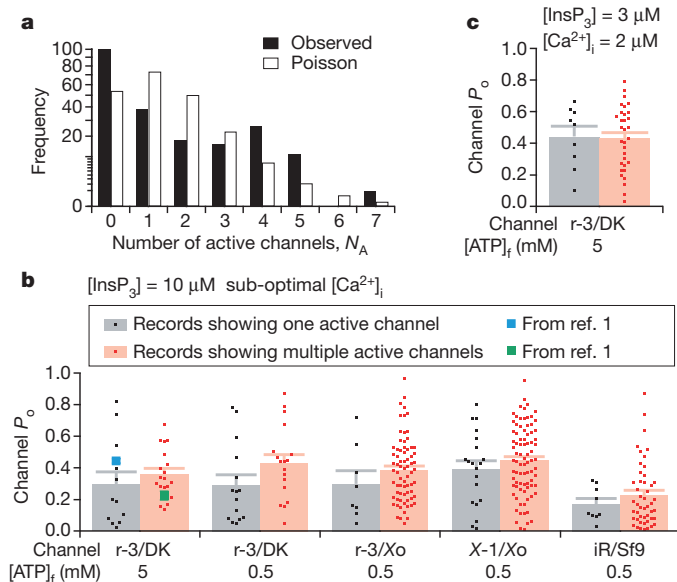


Figure 1 | InsP₃R channels are clustered before exposure to InsP₃, with gating properties unaltered by clustering. **a**, N_A in nuclear membrane patches with no pre-exposure to InsP₃ obtained from InsP₃R3 expressing DT40-KO cells. Note nonlinear square-root scale for frequency axis. **b**, P_o observed under saturating [InsP₃] and sub-optimal [Ca²⁺]_i in multi- and apparent single-channel current records for recombinant rat InsP₃R3 (r-3) channels expressed in DT40-KO (DK) cells or *Xenopus* oocytes (Xo), endogenous *Xenopus* InsP₃R1 (X-1) channels from *Xenopus* oocytes (Xo), and endogenous insect InsP₃R (iR) channels from Sf9 cells. Concentrations of free ATP⁴⁻ ([ATP]_i) in the pipette solutions used are indicated. Mean P_o with s.e.m. (as error bars) and P_o for individual current records are shown, together with mean P_o from Rahman *et al.*¹ **c**, P_o of r-3 channels in DK cells in optimal [Ca²⁺]_i and sub-saturating [InsP₃], ligand conditions not investigated in Rahman *et al.*¹. Same symbols as in **b** are used.

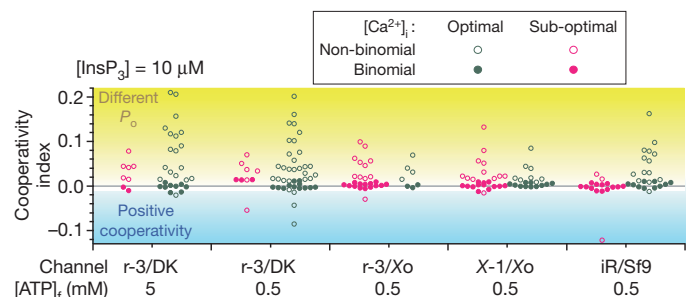


Figure 2 | Distribution of cooperativity index for two-channel current records of different InsP₃R channels in various systems in optimal and sub-optimal [Ca²⁺]_i. Filled and open circles represent records with two channels exhibiting identical and independent, or non-binomial gating, respectively. Non-binomial records with cooperativity index, $(P_2 + P_1/2) - P_2$, significantly greater than 0 (in yellow shaded region) had two channels gating with different P_o , and those with cooperativity index significantly smaller than 0 (in blue shaded region) had two channels gating with positive cooperativity. The cooperativity indices have no correlation with the durations of the current records (data not shown) and therefore are unlikely to be significantly affected by current record durations limited by channel inactivation.

which all but one channel inactivated before gigaohm seal formation. Yet, the mean number of active channels (N_A) we observed for InsP₃R3 in DT40 nuclear patches (1.36 ± 0.12 for 211 patches) is similar to that reported in Rahman *et al.*¹, suggesting that inactivation did not substantially impair our ability to count channels in these patches. We did detect larger N_A (10.8 ± 1)⁸ in outside-out nuclear patches, which probably have significantly larger membrane areas and were isolated using a different technique, and therefore are not an appropriate comparison to illustrate either the variability in InsP₃R expression level in DT40-KO cells or the effect of inactivation on N_A detected.

The P_o distributions that we observed in apparent single-channel and true multi-channel patches were similar, with no indication that two populations of channels with different P_o exist. Furthermore, mean P_o of our true multi-channel patches is comparable to the lone-channel P_o observed by Rahman *et al.*¹ (Fig. 1b). Thus, our conclusion that clustering does not affect InsP₃R channel gating is not compromised by the irreversible inactivation of InsP₃R channels.

We have no clear explanation for the discrepancies between our observations and those reported by Rahman *et al.*¹. However, neither InsP₃-induced InsP₃R clustering nor its modification of InsP₃R gating are consistently observed for InsP₃R expressed in DT40-KO and other cells. In contrast, channel clustering before InsP₃ exposure was observed in all cell systems investigated without effect on channel gating^{3–5}. Thus, we suggest that InsP₃-induced channel clustering and modification of channel gating by clustering may not be universal phenomena.

METHODS SUMMARY

Single-channel P_o for a sufficiently long⁵ current record with N_A active channels was evaluated as $= \sum_{i=0}^{N_A} [iP_i] / N_A$, where P_i , the probability that i channels were active simultaneously in the record, was determined by the same method as Rahman *et al.*¹. The channel gating pattern for a two-channel current record was determined from P_i . If P_i are similar to the expected binomial values, $[2! P_o^i (1 - P_o)^{2-i}] / [i! (2-i)!]$ for $i=0, 1, 2$ ($P > 0.05$ by χ^2 -test), the channels gated independently with similar P_o . Otherwise, if the cooperativity index,

$(P_2 + P_1/2)^2 - P_2$, is >0 , they gated with different P_o , with or without negative cooperativity. If $(P_2 + P_1/2)^2 - P_2 < 0$, they gated with positive cooperativity⁹.

Horia Vais¹, J. Kevin Foskett^{1,2} & Don-On Daniel Mak¹

¹Department of Physiology, University of Pennsylvania Perleman School of Medicine, Philadelphia, Pennsylvania 19104, USA.

e-mail: dmak@mail.med.upenn.edu

²Department of Cell and Developmental Biology, University of Pennsylvania Perleman School of Medicine, Philadelphia, Pennsylvania 19104, USA.

Received 20 October 2010; accepted 11 August 2011.

1. Rahman, T. U., Skupin, A., Falcke, M. & Taylor, C. W. Clustering of InsP₃ receptors by InsP₃ retunes their regulation by InsP₃ and Ca²⁺. *Nature* **458**, 655–659 (2009).
2. Rahman, T. & Taylor, C. W. Dynamic regulation of IP₃ receptor clustering and activity by IP₃. *Channels (Austin)* **3**, 226–232 (2009).
3. Mak, D.-O. D. & Foskett, J. K. Single-channel kinetics, inactivation, and spatial distribution of inositol trisphosphate (IP₃) receptors in *Xenopus* oocyte nucleus. *J. Gen. Physiol.* **109**, 571–587 (1997).
4. Mak, D.-O. D. *et al.* Single-channel properties in endoplasmic reticulum membrane of recombinant type 3 inositol trisphosphate receptor. *J. Gen. Physiol.* **115**, 241–256 (2000).
5. Ionescu, L. *et al.* Graded recruitment and inactivation of single InsP₃ receptor Ca²⁺-release channels: implications for quantal Ca²⁺ release. *J. Physiol. (Lond.)* **573**, 645–662 (2006).
6. Mak, D.-O. D., McBride, S. & Foskett, J. K. Regulation by Ca²⁺ and inositol 1,4,5-trisphosphate (InsP₃) of single recombinant type 3 InsP₃ receptor channels. Ca²⁺ activation uniquely distinguishes types 1 and 3 InsP₃ receptors. *J. Gen. Physiol.* **117**, 435–446 (2001).
7. Mak, D.-O. D., McBride, S. & Foskett, J. K. Inositol 1,4,5-trisphosphate activation of inositol trisphosphate receptor Ca²⁺ channel by ligand tuning of Ca²⁺ inhibition. *Proc. Natl Acad. Sci. USA* **95**, 15821–15825 (1998).
8. Vais, H. *et al.* Redox-regulated heterogeneous thresholds for ligand recruitment among InsP₃R Ca²⁺ release channels. *Biophys. J.* **99**, 407–416 (2010).
9. Kenyon, J. L. & Bauer, R. J. Amplitude histograms can identify positively but not negatively coupled channels. *J. Neurosci. Methods* **96**, 105–111 (2000).

Author Contributions All authors contributed to project planning and wrote the paper. H.V. and D.-O.D.M. did data acquisition and analysis.

Competing financial interests: declared none.

doi:10.1038/nature10493

Rahman *et al.* reply

REPLYING TO H. Vais, J. K. Foskett & D.-O. D. Mak *Nature* **478**, doi:10.1038/nature10493 (2011)

We observe that under conditions designed to mimic those of an unstimulated cell, InsP₃ causes InsP₃R to assemble rapidly and reversibly into small clusters of about 4–5 channels, and that gating of lone and clustered InsP₃R by InsP₃ and Ca²⁺ is different. We speculate that dynamic assembly of InsP₃R clusters by InsP₃ may contribute to the genesis of elementary Ca²⁺ release events by both assembling clusters and re-tuning the behaviour of InsP₃R within them^{1–3}. Vais *et al.*⁴ report that they have been unable to replicate some of our findings.

We cannot explain their results, but we note the considerably greater variability in their observations (Fig. 1a). This may be related to difficulties in reliably estimating the number of InsP₃R (N_A) within a patch. They reported that under the conditions used in their figure 1 (ref. 4), “abrupt, irreversible channel inactivation was consistently observed”⁵ in Sf9 cells^{5,6}, DT40 cells^{5,7} and *Xenopus* oocytes⁷, with time constants of ~30–120 s (ref. 7). We observe no inactivation¹. This loss of activity, before and after seal formation⁶, may cause unreliable estimation of N_A . Indeed, in patches isolated from the same DT40 nuclei, they report N_A of 1.36 for lumen-out patches⁴ but 10.8 for cytosol-out patches where inactivation is much slower⁵. Results in their

figure 1b (ref. 4) show some patches in which open probability (P_o) is <0.02 . Recordings (>65 s) comparable to the time course of InsP₃R inactivation⁴ are required to establish confidently that such patches contain only a single channel^{1,6}.

Vais *et al.* also suggest elsewhere that N_A increases with the intensity of InsP₃R stimulation^{5,6}, and that the effect is modulated by sulphhydryl reagents⁵. Our limited analysis failed to replicate this³, but it is noteworthy that our analyses¹ and theirs⁴ differ in the amounts of 2-mercaptoethanol included in culture media. Considerable variation in the amplitude of InsP₃R3 unitary currents^{5,8} and the frequent occurrence of sub-conducting states within records⁸ may also confound their estimates of N_A . Counting InsP₃R (figure 1a in ref. 4), calculating P_o (figure 1b, c in ref. 4), and assessing whether P_o is binomially distributed (figure 2 in ref. 4) depend on reliably determining N_A . Our recordings are stable for many minutes with no inactivation or InsP₃R recruitment, and only extremely rare visits to sub-conducting states^{1,2}. We suggest that the more complex behaviour of InsP₃R observed by Vais *et al.*⁴ may compromise estimates of N_A , so that even patches in which they detect only one active InsP₃R may

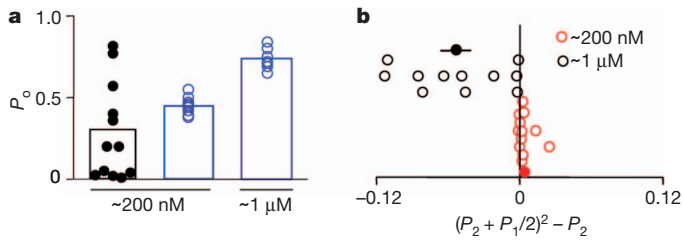


Figure 1 | InsP₃-evoked clustering of InsP₃R alters their gating. **a**, InsP₃R3 expressed in DT40-KO cells were stimulated with 10 μ M InsP₃ in cytosol-like medium containing a free $[Ca^{2+}]$ of ~ 200 nM or ~ 1 μ M. P_o is shown for patches with a single InsP₃R (blue) taken from our results¹ and for equivalent analyses taken from figure 1b of Vais *et al.*⁴ (black). **b**, The analysis reported in ref. 4 was used to calculate the probabilities of 1 (P_1) or 2 (P_2) InsP₃R being open in patches with 2 InsP₃R after stimulation with 10 μ M InsP₃ and the indicated concentrations of free $[Ca^{2+}]$. We applied their analysis⁴ to determine whether InsP₃R gated independently or cooperatively by computing $(P_2 + P_1/2)^2 - P_2$, where $(P_2 + P_1/2)$ provides the estimate of P_o . Results are shown for each of 12 and 11 patches with 200 nM and 1 μ M free $[Ca^{2+}]$, respectively (open symbols); and as mean \pm s.e.m. (solid symbols: 0.0048 ± 0.0011 and -0.0517 ± 0.0127 , for 200 nM and 1 μ M $[Ca^{2+}]$, respectively; error bars are smaller than the symbol for the former). The results are consistent with independent gating of InsP₃R in 200 nM free $[Ca^{2+}]$, and with positively cooperative gating in 1 μ M free $[Ca^{2+}]$.

contain several inactivated InsP₃R that may cluster and regulate a neighbour¹.

Finally, and to allow direct comparison with the analysis Vais *et al.* show in figure 2 of ref. 4, we re-analysed our data¹ and included results from InsP₃R1 (see Fig. 1b). This provides statistical evidence that in patches with two InsP₃R, InsP₃R gate independently at resting $[Ca^{2+}]$ but with positive cooperativity at elevated $[Ca^{2+}]$.

Taufiq Rahman¹, Alexander Skupin^{2,3}, Martin Falcke⁴ & Colin W. Taylor¹

¹Department of Pharmacology, Tennis Court Road, Cambridge CB2 1PD, UK.

e-mail: cwt1000@cam.ac.uk

²Luxembourg Centre for Systems Biomedicine, 162a Avenue de la Faïencerie, University of Luxembourg, L-1511, Luxembourg.

³Institute for Systems Biology, 401 Terry Avenue North, Seattle, Washington 98103, USA.

⁴Mathematical Cell Physiology, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str, 10, 13092 Berlin, Germany.

1. Rahman, T. U., Skupin, A., Falcke, M. & Taylor, C. W. Clustering of InsP₃ receptors by InsP₃ retunes their regulation by InsP₃ and Ca^{2+} . *Nature* **458**, 655–659 (2009).
2. Rahman, T. & Taylor, C. W. Dynamic regulation of IP₃ receptor clustering and activity by IP₃. *Channels* **3**, 226–232 (2009).
3. Taylor, C. W. & Tovey, S. C. IP₃ receptors: toward understanding their activation. *Cold Spring Harb. Perspect. Biol.* **2**, a004010 (2010).
4. Vais, H., Foskett, J. K. & Mak, D.-O. D. InsP₃R channel gating altered by clustering? *Nature* **478**, doi:10.1038/nature10493 (2011).
5. Vais, H. *et al.* Redox-regulated heterogeneous thresholds for ligand recruitment among InsP₃R Ca^{2+} -release channels. *Biophys. J.* **99**, 407–416 (2010).
6. Ionescu, L. *et al.* Graded recruitment and inactivation of single InsP₃ receptor Ca^{2+} -release channels: implications for quantal Ca^{2+} release. *J. Physiol. (Lond.)* **573**, 645–662 (2006).
7. Mak, D.-O. D., White, C., Ionescu, L. & Foskett, J. K. in *Calcium Signaling* (ed. Putney, J. W. Jr) 203–229 (Taylor and Francis, 2006).
8. Vais, H., Foskett, J. K. & Mak, D. O. Unitary Ca^{2+} current through recombinant type 3 InsP₃ receptor channels under physiological ionic conditions. *J. Gen. Physiol.* **136**, 687–700 (2010).

Author Contributions T.R. performed the experiments and, with C.W.T., analysed the data. C.W.T. and T.R. wrote the paper with input from A.S. and M.F. The project was directed by C.W.T.

doi:10.1038/nature10494

natureOUTLOOK

MEDICAL RESEARCH MASTERCLASS

13 October 2011 / Vol 478 / Issue No. 7368



COVER ART: NIK SPENCER

Editorial

Herb Brody,
Michelle Grayson,
Tony Scully

Art & Design

Wes Fernandes,
Alisdair Macdonald,
Kate Duncan

Production

Karl Smart, Emilia
Orviss, Leonora
Dawson-Bowling,
Stephen Russell

Sponsorship

Gerard Preston, Reya
Silao, Yvette Smith

Marketing

Elena Woodstock,
Hannah Phipps

Project Manager

Claudia Deasy

Art Director

Kelly Buckheit Krause

Magazine Editor

Tim Appenzeller

Editor-in-Chief

Philip Campbell

Sure, the beautiful surroundings of Lake Constance are both blissful and stimulating. But what really defines the annual meeting of Nobel laureates on the German island of Lindau is the interaction between the 600-or-so young researchers and the score of Nobel laureates. The respected elders of the meeting mingle freely with the students — and laureates and students appear equally inspired by each other. It is this interaction that we aim to capture in this *Nature Outlook*. Students sat next to laureates as they listened to a series of presentations that ranged from personal recollection, to scientific discourse, to advice on career development. Laureate Elizabeth Blackburn opened the lecture series in 2011 with an ebullient talk about her Nobel-winning research into telomeres. These protective chromosomal end caps are still yielding secrets (page S13).

The soul of the Lindau meetings is in the casual yet formative interactions among all the scientists. This year there was a new format: the masterclass, in which a laureate picks a handful of young researchers to present their work in front of a large audience. In our report, we describe how Aaron Ciechanover guided and engaged his students in a provocative discussion that challenged everyone present to broaden their thoughts (page S19).

We have a series of Q&As with some of the laureates who attended the meeting this year (pages S2 to S12) with some questions posed by young scientists at lindau.nature.com.

The United States was the sponsoring country this year, and we acknowledge that support by examining the promise and perils of US attempts to turn research into knowledge, drugs and devices that promote patient care and health (page S16) — topics that were enthusiastically discussed at Lindau.

We are pleased to acknowledge the financial support of Mars, Incorporated and Porsche in producing this *Outlook*. As always, *Nature* retains sole responsibility for all editorial content.

Michelle Grayson

Associate Editor, Nature Outlook.

CONTENTS

- S2 THOMAS ARTHUR STEITZ**
Lunchtime science
- S4 AARON CIECHANOVER**
On the wings of imagination
- S5 EDMOND HENRI FISCHER**
Biochemistry without boundaries
- S6 ADA ETIL YONATH**
Tough science
- S8 JEAN-MARIE LEHN**
Rational enthusiasm
- S10 FERID MURAD**
Heart of science
- S12 HARALD ZUR HAUSEN**
Science addict
- S13 TELOMERES**
All's well that ends well
Could tinkering with the tips of chromosomes help treat disease and even wind back our biological clocks?
- S16 TRANSLATIONAL RESEARCH**
The American way
How the United States is performing in turning its world class science into new drugs and diagnostics
- S19 MASTERCLASS**
The gathering brainstorm
The work of three young researchers is scrutinized by a laureate, and all are encouraged to think outside the box

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf.

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol XXX, No. XXXX Suppl. Sxx–Sxx (2011).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Medical Research Masterclass* supplement can be found at <http://www.nature.com/nature/outlook/masterclass2011/>. It features all newly commissioned content as well as a selection of relevant previously published material.

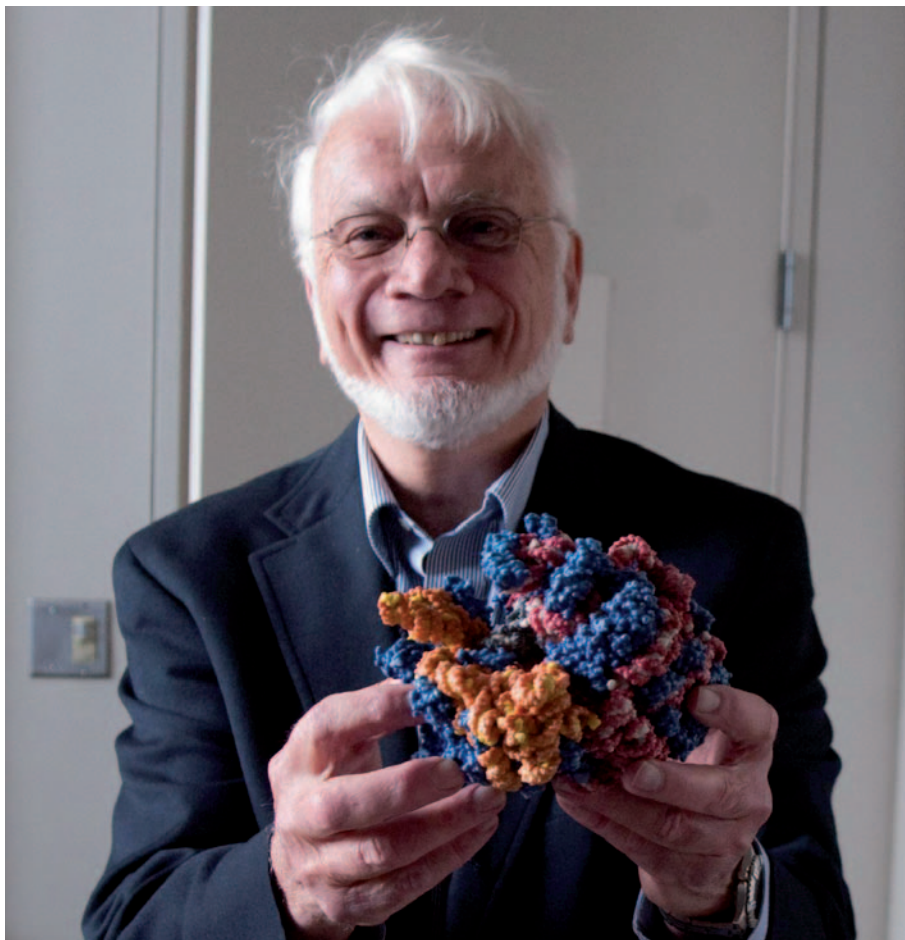
All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas — including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

Feedback@nature.com
Copyright © 2011 Nature Publishing Group



Thomas Arthur Steitz

Lunchtime science

Biophysicist at Yale University in New Haven, Connecticut. Shared 2009 Nobel Prize in Chemistry for knowledge of the structure and function of the ribosome — the intracellular machine that builds proteins from instructions carried by RNA. Born in Milwaukee, Wisconsin in 1940. The oldest of five children, Steitz has admitted to being an average student in high school, until motivated to compete against his youngest brother who was getting better grades. Steitz was a keen musician and chorister, and considered a career in music before finally choosing to pursue science.

In Lindau, you said that eating lunch alone in your office is bad for doing good science — do you always think and act scientifically?

The reason it's important to have lunch with colleagues, students, postdocs, faculty etc., is so you can talk about ideas and experiments and science. It's a great opportunity to connect with others. See what they are doing, tell them what you're doing.

I picked this up from my years at the Laboratory of Molecular Biology in Cambridge, UK. Everybody would have coffee in the morning, lunch and tea together in the afternoon. They would always gather around tables and

exchange ideas. It helps to stimulate thinking, to give ideas about new experiments; or you might realize that an experiment you wanted to do is perhaps not the best idea. It goes both ways.

Here [at Yale University] we often have lunch with faculty from other departments, such as geology, chemistry or physics. Obviously you can't talk about your experiments in as much detail with these colleagues, but I have still learned a lot of interesting things, for example about global warming — erosion of salt marshes and rising sea levels.

The view that scientists are insular, quiet

and uncommunicative is not correct. I've had some students like that in my lab, but they're not the best scientists. I can't exclude the possibility that it's field dependent: theoretical physics might be one scientific discipline where it's more important to think through your own thoughts. But certainly in biological sciences, which are so complex with lots of bits of information to piece together, having conversations and learning things is very useful in solving puzzles.

Has winning the Nobel prize changed the way you eat lunch?

I'm not able to eat with colleagues as often as previously. But then again I'm also able to have lunch in many more places in the world, such as China.

What most motivated you in your work: the Nobel prize, being published, enthusiasm, curiosity or a desire to be of service to mankind?

I was never motivated by the Nobel prize; I think that's the wrong goal. I'm just curious, and everybody I know who does well in science is very curious. You have questions and you want to know the answers, and you get excited when you find things out. You are the kind of person who wants to turn over the rock and see what's underneath. This curiosity is also driven by the adrenaline rush that hits you when you find the answer to a question you have been asking.

Of course, when it comes to being of service to mankind — this is important when considering which questions to ask in the first place. Some problems have more practical impact than others. For example, in the late 1980s, several postdocs came into my office saying that we should do something about HIV, as it was becoming clear it was a big problem. I decided that we should work on HIV reverse transcriptase. That made sense to us because we had been studying DNA polymerases involved in replication, so we could then ask about the differences between reverse transcriptase and DNA polymerase — and, by the way, we'll study it in complex with a non-nucleotide inhibitor so we can also help with drug design. That way we could kill two birds with one stone: furthering our research into these types of enzymes and also into HIV proteins that can assist in developing drugs against it. So it's a combination of following your own nose and doing something that will be helpful.

We also solved the first structure of synthetase and transfer RNA (tRNA) in 1989. We went on to solve the structure of synthetase and tRNA in complex with an antibiotic to see if we could design new antibiotics. You have to look for opportunities that achieve a number of goals.

My major goal is not to design pharmaceuticals, important and exciting as that is to people in the pharmaceutical industry, but I'm not opposed to doing that if it fits in with other objectives.

NATURE.COM
For some of the latest research on the ribosome:
go.nature.com/5al5ge



Steitz's lecture at the Lindau meeting in 2011 discussed a broad range of topics from the structure of the ribosome to the better design of antibiotics.

As a recent laureate, what impact do you hope your Nobel prize will have?

I frankly think that just because I'm good at structural biology and uncovering the mysteries of how molecules work, I'm no expert on peace. I have opinions, but they should not be any more highly regarded than any others.

That said, people pay more attention to what laureates say — more than they should — and you should take advantage of the opportunity when it comes. However, that's not my objective at the moment. I get letters from various organizations wanting us to sign letters for this or against that, and I'll sign if I agree with it. But in terms of going on a campaign — that's not on my list of things to do.

Do you think bacterial antibiotic resistance is as big a threat as the media makes out?

I think it's gargantuan. About 100,000 people die each year of methicillin-resistant

Staphylococcus aureus (MRSA). More people die of MRSA infection than from AIDS. It is very important to have more antibacterial drug development.

Fungal infections are even more difficult to treat. Bacteria are prokaryotes, but fungi are eukaryotes like us, so we need drugs that can target the fungus but not the host.

Together with a few colleagues, I founded Rib-X Pharmaceuticals about ten years ago. The idea was to use the structure of the bacterial 50S ribosomal subunit in complex with existing antibiotics, to see where they bind and what the reactions are, in order to design new antibiotics. We use computational chemistry (initially from Jorgensen's Lab at Yale) to design new compounds that will be effective against bacteria. One compound (radezolid, a next-generation oxazolidinone) is through phase II clinical trials and appears to be effective against all strains of MRSA.

Another set of compounds are designed completely de novo, based on knowledge of the target site. These compounds appear to be effective against both Gram-negative and Gram-positive bacteria.

Will medicines move away from being chemical to being more biological?

I think biological medicines are useful in many diseases, but not as antibacterials. I don't see anything on the horizon. There is the challenge of delivering the drug: it has to be stable and soluble in the host skin or bloodstream. And then it has to be able to get inside bacterial cells. Nucleic acids are not easy to get in — or cheap.

Antibacterials are really part of a microbial biological warfare. Bacteria have been fighting with each other for millions of years; they use these chemicals to try to kill other species. That is where antibiotics initially came from, they were all isolated from species of bacteria or fungi. Nature figures out how to do it, and we're just following nature.

Will we be fighting this war forever?

Bacteria have learnt to overcome everything we have thrown at them so far. It goes to show that evolution trumps intelligent design. ■



I would prefer if Nobel prizewinners said that they were most motivated by observing the suffering of mankind, or motivated by love and kindness. But I am really happy that they have given an honest response.

K.L. Senarath Dayathilake, psychologist, Human Well-being Science Program, Kotte, Sri Lanka, who posed the original question on lindau.nature.com.



Aaron Ciechanover

On the wings of imagination

Biochemist at Technion, the Israeli Institute of Technology in Haifa. Shared the 2004 Nobel Prize in Chemistry for the discovery of the ubiquitin system, which mediates protein degradation in all plant and animal cells by destroying proteins that are denatured, misfolded or no longer needed. Family moved from Poland in the 1920s, and he was born in Haifa in 1947. The following year the state of Israel was established.

You won a share of the Nobel prize for your discovery of the cell's protein degradation process. How has the field changed since those early days?

The field has changed dramatically. The ubiquitin proteolytic system is now known to regulate many basic cellular processes, such as cell division, differentiation, transcription and quality control. Altogether, components of the system comprise 6–7% of all genes in the human genome: well above what was expected for just a 'protein scavenger system'. Consequently, it is no surprise to find that aberrations in this system are implicated in the pathogenesis of many diseases. This knowledge has led to the development of mechanism-based drugs, including one on the market to combat multiple myeloma. It is very exciting. When we started we were not looking for the ubiquitin system; we were after an answer to the question of how proteins are degraded. We had no idea that such a complex system was behind it.

How optimistic are you that diseases of

protein accumulation, such as Alzheimer's disease, will be treatable?

There are diseases that involve gain of function and diseases that involve loss of function. With the former, one can be more optimistic because the solution is to develop inhibitors and antagonists, which are easier to develop than agonists or stimulants. In pharmacology, it is easier to slow down a system than to speed it up, so it depends which side of the ubiquitin system you are talking about. On the cancer side — as it involves a gain of function — progress will come faster; on the neurodegenerative side, such as Alzheimer's or Parkinson's disease, it may take longer.



It is interesting that our preparations for combating complex diseases such as Alzheimer's and cancer are described as 'primitive'. What are the hurdles for progress — is it a lack of funding or are there knowledge gaps? Also, what attitude and skill set do young researchers need to tackle such complexities of multidisciplinary research?

G. N. Viswanath, professor of management, Bharatiya Vidya Bhavan, Bangalore, India, who posed the original question on lindau.nature.com.

➔ What are the limitations of modern medicine and how can they be overcome?

Modern medicine is rather primitive. We can only detect diseases like cancer and neurodegenerative disorders when they are well advanced; the challenge is to diagnose them much earlier, or even to predict and prevent them. We also tend to group different diseases under a single umbrella, for example grouping prostate or breast cancers together and treating them similarly. The fact the treatment succeeds in some people yet fails in others suggests that these groupings are erroneous and each disease is comprised of distinct sub-groups. With personalized medicine, we will be able to profile individual patients more precisely. Such profiling will be based initially on the patient's genome, but later it will use the transcriptome, proteome and metabolome. The road is long; the technologies and, crucially, the interpretation of results are not there yet, but the future is bright. The power of personalized medicine will reside not only in our ability to develop new drugs to currently incurable diseases, but also to predict, and therefore to prevent diseases based on individual patients' susceptibility. Along the way it will be important to pay special attention to bioethical issues concerning exposure of patients' most intimate information.

What are the challenges of translational medicine?

There are two potential approaches: carry out curiosity- and hypothesis-driven research that flies on the imagination of brilliant scientists, or carry out disease- and drug-oriented research. Curiosity-driven research in the last century brought tremendous development in biomedicine — novel drugs and sophisticated devices — and I strongly believe this is the way to go. If we shift mostly to translational research, the springs of knowledge will dry up, and there will be nothing left to translate. Perhaps the public and our political leaders think that things are going too slowly, but that's a dangerous perspective.

Do you have any advice for graduate students trying to pick a research topic?

Choose a good mentor who asks original questions. Be patient, do not give up: work hard and persevere. Be passionate and excited about what you are doing; think of your scientific profession as if it were your hobby. Luck is important too, but remember, very often luck is not blind: it hits those who are ready. ■



Edmond Henri Fischer

Biochemistry without boundaries

Biochemist at the University of Washington in Seattle, he won a share of the 1992 Nobel Prize in Physiology or Medicine for discoveries concerning reversible phosphorylation: a regulatory mechanism that activates and deactivates enzymes in the vast majority of living cells. Fischer was born in Shanghai, China, in 1920.

Do you keep abreast of advances in protein phosphorylation? What excites you these days?

Protein phosphorylation is involved in a number of diseases directly, including: Alzheimer's disease; Parkinson's disease; diabetes; myelogenous leukaemia; viral and bacterial infections such as smallpox, cholera and plague; and cancer. A lot of biotechnology and pharmaceutical companies are working on protein phosphorylation — on enzymes that add phosphates (kinases) and those that take them off (phosphatases). Too little work on phosphatases in my opinion, particularly as certain ones, like PTEN, act as tumour suppressors. However, very little has come out yet regarding drugs that target phosphatases.

➔ A lot of science is quite reductionist. Is this better than taking a systematic, top-down view?

No, one of the beauties of science is that you never know when the next big breakthrough will happen. You approach it systematically. Science builds on science. Every result pertains to the next questions, and every question suggests the next experiment. You cannot predict when the next big breakthrough will happen.

Research fields emerge unpredictably. Take protein phosphorylation — that was our luck. We found a very simple reaction — embarrassingly simple — which turned out to be absolutely

crucial for the regulation of cellular processes. For years it was considered to be the most prevalent mechanism of cellular regulation, and then all of a sudden ubiquitination popped up (co-discovered by Nobel laureate Aaron Ciechanover; see page S4) and now that has taken the centre stage.

In the 1950s it was all about enzymes. We had little information on enzymes or on the 3-dimensional structure of proteins until Max Perutz and John Kendrew (recipients of the 1962 Nobel Prize in Chemistry) used X-ray crystallography to determine the structures of myoglobin and haemoglobin.

Has working on proteins for so long given you a particular view on life?

Life is an inevitable phenomenon and it exists all over the universe. There is very good evidence that there are 500 million planets similar to Earth in the Milky Way. The probability that life does not exist on one of these other planets approaches

zero. What kind of life, nobody knows.

A self-replicating system was established 3.5 billion years ago. At first, evolution was slow, becoming more rapid in the past 550 million years as single-cell organisms coalesced to form multicellular organisms. Before that, single cells had to compete with one another for food, light, micronutrients, etc. But the moment cells began to associate with one another to form specific tissues and organs, they had to cooperate: to speak with one another, synchronize growth and behaviour for the good of the whole.

The main criteria for life are self-duplication and conversion of foreign molecules into molecules identical to you. So crystallization is not life. Putting a crystal of copper sulphate in solution so that it grows is not life. Some people think the first self-duplicating molecule was a protein, but other people think it might be nucleic acid.

Your Nobel prize is for physiology or medicine. How do these areas relate to each other?

Medicine is not only clinical, it also covers molecular biology, biochemistry, physiology, pathology, neurobiology and more. These days, medicine extends from the molecule to the organism.

To most researchers, it doesn't make any difference if they are in the department of biochemistry or the department of neurobiology or pathology or microbiology. The amount of new information that is being gathered in the biomedicine arena is such that it is almost inconceivable that people could work in isolation. Collaboration is totally indispensable if one wants medicine and biology to progress. We have all the tools; people communicate very easily now.

If you compare what you had planned at the very beginning of your career, and where you are right now, are they similar?

When I was a kid I didn't want to be a biochemist, I wanted to be a microbiologist. When I started at the University of Geneva in Switzerland I sought advice from the professor of microbiology. He said if I wanted to be a research biologist, I should study chemistry first. "One uses test tubes more than microscopes these days," he said. So I became an organic chemist, but always with one eye on living systems. That is why my thesis was on enzymes. I have always been interested in basic research. The only thing that has changed is that I used to lack the systematic approach needed to solve problems. ■



Fischer's response implies that new scientific discoveries affect the ontological paradigms in play, and that these vary. I wonder: does this suggest that what is considered a fundamental scale at one point of investigation becomes superseded by a more complex set of things in the next?

Amanda Parr, former philosophy student and now a public servant in Canada, who posed the original question on lindau.nature.com.

➔ **NATURE.COM**
Nature Video of
Fischer advising a
young researcher:
go.nature.com/a5eijf



Ada Etil Yonath

Tough science

X-ray crystallographer currently at the Weizmann Institute of Science in Rehovot, Israel. She won a share of the 2009 Nobel Prize in Chemistry for her work on the structure and function of the ribosome. Yonath was born in 1939 in Jerusalem to a poor family. Her father died when she was 11 years old, and Yonath helped support her mother and younger sister. Yonath was the first Israeli woman to win a Nobel prize and the first woman in 45 years to win the Nobel Prize in Chemistry

In your talk at Lindau, you used a lot of analogies and evidence from outside of your scientific work — climbing mountains, or finding inspiration in polar bears. Do you try to keep a broad mind and outside interests?

I don't make any special effort to look for inspiration. I just happened to read about polar bears because I had time to read after having an accident and had to recover. My mind is usually wandering around — I read one thing and think about related subjects. My ideas are always bubbling under the surface and inspiration can strike from any source.

You have spoken openly about your humble beginnings: do you think they helped you develop your inquisitive mind?

I think my difficulties as a child — the poverty and fight for survival — made me stronger. I didn't even know what science was as a child, but I have found that science is a blessing. I can come up with a question and be paid to answer it. Of course the question has to be fundamental and I have to prove that I might be able to do it and show that I'm making progress. In basic science, many problems I come across, even if I can't solve them, I can find a way around them or at least take a different approach to the problems. In life, if you

can't help your mother to feed you and your sisters, there are no alternatives. Survival in life gave me a lot of persistence and resistance to keep going and actually solve my problems. And it helped me put things in perspective.

Did these qualities of persistence and resistance help you determine the structure of the ribosome?

I try to emphasize that science is a step-by-step process of making small advances. Not everyone understands that. It's not like one day there was a miracle. In our work, we might have had misgivings, but we also had hints of progress. Each step was harder and harder and brought

"I'm trying to change the image of scientists, especially of female scientists, so there will not be so many anti-female sentiments."

more problems, but overcoming them was very beneficial, and not just for our work. When my project on the ribosome was in its initial phase, I was pushing the management committee for X-ray synchrotrons to be technically more suitable so that even if we didn't solve the

structure of the ribosome, we would at least contribute a lot in terms of improving the technologies available at synchrotrons, both for us and for other scientists.

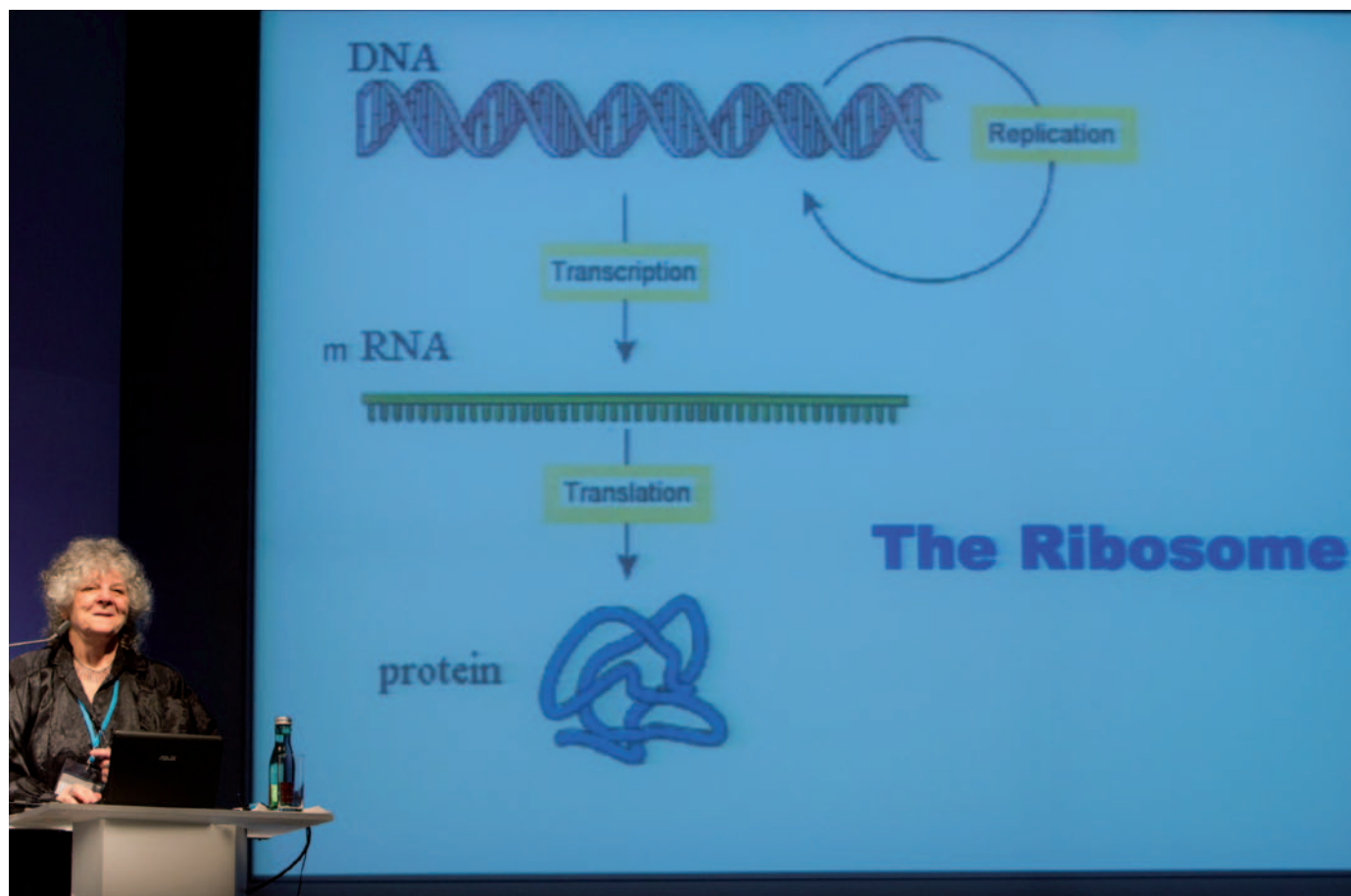
The complex ribosome structure was a great challenge to crystallize. Is there any limit on the type of structure that can be crystallized?

The ribosome is one of the most difficult structures to have been crystallized. Size isn't the biggest problem in crystallization — although it used to be — it's internal mobility and flexibility: heterogeneity. A ribosome is a complex of about 80 different components that vary according to functional state. But when you do crystallization studies you require one structure. If you have a population of ribosomes that are all in different configurations you will get different 3-dimensional shapes and no periodicity, which you need in crystallography. These are the challenges that are difficult to overcome.

At Lindau, you spoke of the ribosome's 'ancient core of RNA'. Do you think RNA is more important than DNA in the origins of life?

Yes, definitely: DNA came much later. One aspect of RNA is that it is the thing carrying the information from DNA to be translated into proteins in the ribosome. But RNA is not just an information store as DNA is; in addition to self replication, it can actually do things and make things, such as enzymes to control reactions. DNA cannot do this.

There are many different types of RNA. One type of RNA carries the protein building blocks. And 60% of the ribosome is RNA — the part that decodes the information and makes the proteins.



Yonath giving her lecture at the 2011 Lindau meeting on the ribosome, one of the most complex cellular component to be determined by X-ray crystallography.

This ribosomal core, which we identified, is an example of a ‘machine’ from before life evolved. Life grew around it. RNA was one of the first simple molecules.

As of 2010, you are one of only 40 women (compared to about 500 men) to have won the Nobel prize. Do you perceive extra pressure to be a scientific role model for young women?

No, I think science is gender-independent. Some people might not think so, because there are problems with equal representation of women in science, but that is not because of the nature of science. The reason why there are not as many women comes from the fact that few women take up jobs to lead groups in labs. This can and should be changed.

Having said that, some people contact me because I am the only woman alive with a chemistry Nobel. They insist that I attend their meeting, which does put a lot of pressure on me. For instance I’m currently in Puerto Rico at the IUPAC World Chemistry Congress; I gave a talk this morning and spoke to the students privately afterwards. Next week I will be the guest of the Chilean students (at the National Commission for Science & Technology), where I will give two lectures: a motivational talk and a scientific talk.

All this attention is enjoyable however, and from the responses I’ve had I think the students

enjoy it too. You can’t imagine how excited they are to meet a Nobel laureate. They want to take pictures and get autographs. It is clear that this is their dream, and maybe I got them closer to realizing it.

As a recent laureate, what impact do you hope your Nobel prize will have?

Very little on my day-to-day work: I am continuing as normal. I have some students that maybe applied to work in my lab because of the prize, which is nice — some from countries like Malaysia, Singapore and China. I’m also directing my efforts to the young generation, showing them that science is not a dull profession, that it’s not about being stuck in a lab all day working on a boring project. I’m trying to change the image of scientists, especially of female scientists, so there will not be so many anti-female sentiments. Because there are people in society, particularly parents, who are not necessarily against women studying science but they don’t agree with women holding demanding positions while running a household.

What did you learn from your mentors?

I remember on one occasion, I made a little mistake — really just a typo. My supervisor was annoyed and questioned me about it. I apologized for it, said it was a minor error and explained that I had a lot of deadlines. He

looked at me and said: “You want to stay in science? Then all your life you will be going from deadline to deadline.” He was right.

How do your early career plans compare with your actual career?

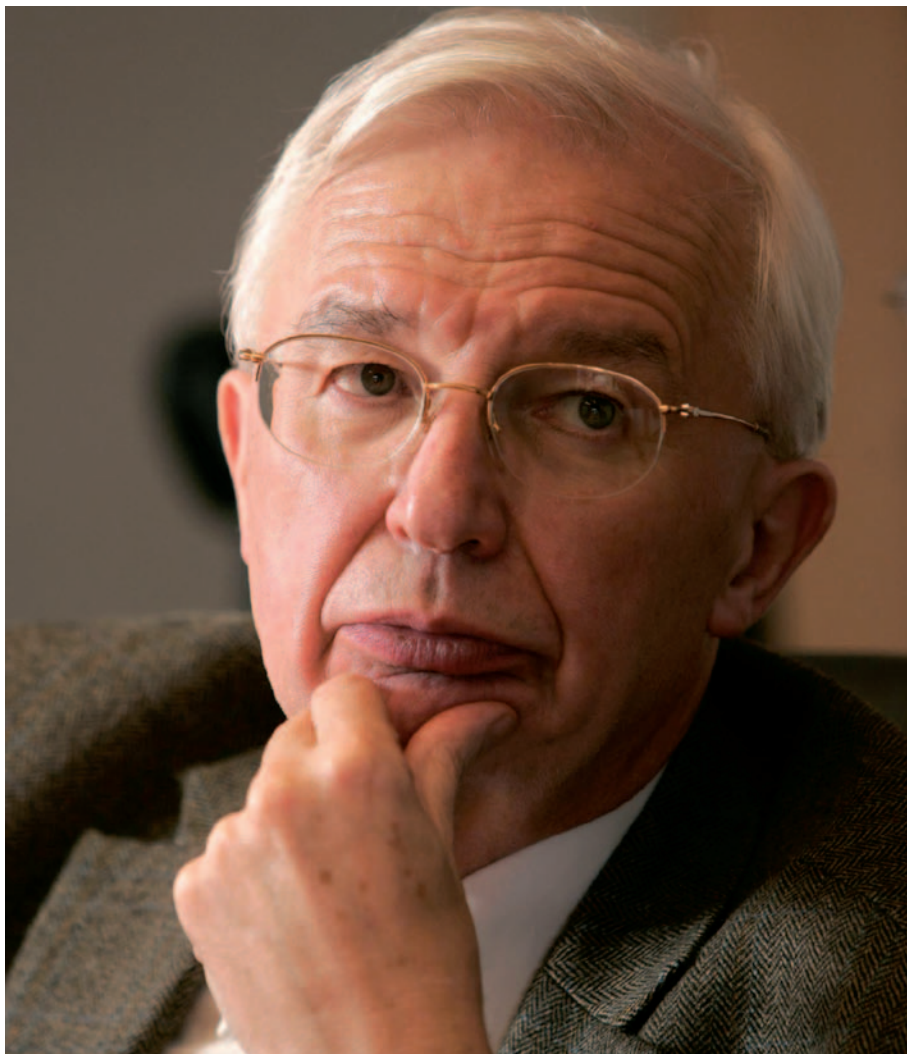
When I was young, I chose something else! My thoughts were more about my country and I didn’t think about science. I did compulsory military service. I served in the medical corps and worked with doctors, planning a course for other doctors in the field. I learned a lot about medicine and doctors in general. Indirectly, this fed into my work in science. I learned not to give up so fast. Try your best to solve the problem; look for all possibilities.

Conversely: how do you try to mentor your students?

I’m just myself. I don’t make any special concessions; I don’t change myself. Most of my students seems to like me as I am.

What kind of attitudes and qualities should a young scientist develop?

Be curious — that is most important. Take passion. Be ready to sometimes experience difficulties, but enjoy what you are doing. In other words: be tough, and love your work. ■



Jean-Marie Lehn

Rational enthusiasm

Chemist at the University of Strasbourg in France. Shared the 1987 Nobel Prize in Chemistry for development and use of molecules that recognize and interact with each other. Coined 'supramolecular chemistry', it is an area of chemistry that exploits non-covalent interactions. Born in 1939 in Rosheim in France, Lehn was the son of a baker who later became the city's organist. Music is Lehn's main passion other than science.

You have spoken of the responsibility of scientists to talk about the role of science in society and culture. Do you think that over the years you have achieved this?

We all try to make what we say attractive, to come across as open and to present things in an understandable way for a broad range of educated people. But it's also important that the people listening make an effort — and that takes time. Some people are receptive and interested. Others have to be hooked in to prepare them to listen. However, a scientist's first responsibilities are to science — knowledge and truth — not

to society. We are here to gain knowledge and understand what's going on, but of course we are also eager to be understood and to benefit society.

Are there any parallels of supramolecular chemistry in society?

There are parallels inasmuch as sociology is the study of humans in society, and supramolecular chemistry is the study of molecules and how they bind to each other, whether that's in love or hate.

For the general public, the concept of 'molecular sociology' can be better understood and

"We would like to be able to analyse and be rational and try not to react to irrational inputs. But it's not easy."

remembered. Molecules can be isolated, like an individual. Our environment can be society or the wider world, but molecules bathe in a world of other molecules: colliding and interacting. Of course, the behaviour of an entity like a human being is much more complex than molecules getting together.

A lot of science is quite reductionist. Is this better than taking a systematic, top-down view?

One has to approach it from both sides. Build from the bottom, with complex behaviours analysed from the top. I don't think one should just do only one way.

We are all made of molecules. All we do is the result of a very complex system, but we can in principle deduce from these components what we are able to do. Higher level properties and behaviours emerge from higher levels of complexity.

The example I usually give is taking an isolated molecule of water, which cannot freeze or boil, yet a glass of water can. A property has appeared at a higher level of complexity, which emerges from the fact that molecules interact in a supramolecular way and act together as a system. Thus as complexity increases, new properties and behaviours emerge at each step that cannot be reduced to what is below, but which can be deduced from it.

Biology is based on molecules; it is the highest expression of the molecular world. It is a demonstration that the simple molecular world can generate great complexity through self-organization.

Do you think that such complex life exists on other planets?

It is very probable. There are so many galaxies and planets that chances are there are similar conditions to those on Earth.

The 4% of observable matter that comprise the bricks of our Universe are composed of elements described in the periodic table. These atoms will be the same everywhere, and they will interact in the same way so that the compounds they form will be the same. For example, a carbon-carbon bond will be the same everywhere. Whether the combinations of elements will be the same is another question.

What most motivates you in your work?

Getting published is just one outcome, and not the major one. The journal in which you publish is not so important (sorry *Nature*, *Science*). What is important is that good work is recognized. It might take longer if it is published in a more obscure journal, but it still counts as being

➔ NATURE.COM
the latest research
on supramolecular
chemistry:
go.nature.com/gtxeuz

published. I can cite a number of papers that led to great discoveries but which were published in second-, third- or fourth-rate journals.

My motivation was simply that I was interested in science and research and in gaining new knowledge. All of the sciences are exciting. The great frustration is that one cannot do everything. You have to be selective to have an impact. Everybody has a given amount of energy, if you apply it on a small area you have a big impact. Too broad and the impact will be weaker.

I got interested in chemistry when I realized it is the basis of all matter. I also realized that chemistry enables you to transform objects of one type of matter into another, using a process that is understandable. Chemistry is able to develop methodologies to gain power over matter, and to generate new compounds that didn't exist before. I like to compare that power to a work of art. When an artist makes a sculpture out of stone, the stone doesn't contain the sculpture, it's the artist who creates it. In the same way, by taking the pieces that the Universe gives us, a chemist can express new combinations often with new, unexpected properties.



Lehn holding his lecture on supramolecular chemistry at the 2011 Lindau meeting.

Do you always think and behave scientifically?

We all try, that is our job, but we are not just pure brains, unfortunately. We would like to be able to analyse and be rational and try not to react irrationally to inputs. But it's not easy.

Many people think of science as just being a body of knowledge, but it is also a spirit — a way to approach things. And this is important to convey as it will help humans live together. Finding answers is also connected to the way you approach the question. Sometimes you stumble onto an answer, but you also have to be able to see that the answer is lying there. Just stumbling is not enough.

Should the world's leading scientific organizations be more united for practical policy planning and research funding?

Not totally united into a single organization. Any organization can make mistakes, for instance, it might be that the evaluation of a given project is not done well. So diversity is good — but too much is chaos. Usually the projects that are in the top or bottom 10–20% are easy to identify. Those in the middle are more difficult, so we need several

institutions for funding and evaluation to judge these projects.

Sometimes a project appears to be terrible, but it eventually leads to something fantastic. So there is a risk that collapsing all the institutions into one would miss these outliers. It's good to have diversity; as with journals. You submit a paper, and some referees like it while some do not; you have opposing views. With research, there are common lines, ideas that come up everywhere. It would be a pity if everybody were to crystallize around the same topics.

There are three types of research: discovery work that furthers our understanding; applied

work in which a scientist makes things for a given purpose; and blue sky thinking with little guarantee of an endpoint. Perhaps this blue sky should be at the beginning of bright people's careers or it could happen at a more advanced stage of their career. It would be good to free these individuals' time: not to have to worry about financing and to let their imaginations go. There are some institutes that are going in this direction, including Germany's Max Planck Society and the Institute for Advanced Study in Princeton, New Jersey. They try to give this freedom.

Should scientists take a type of global research oath at the start of their careers, as medical doctors take the Hippocratic oath?

The first aim of scientific research is to increase knowledge for understanding. Knowledge is then available to mankind for use, namely to progress as well as to help prevent disease and suffering. Any knowledge can be misused. I do not see the need for an oath and I would hope that medical doctors would behave in the same responsible way without the Hippocratic oath.

With respect to weapons and the like, if everybody were to take an oath and refuse to conduct research, then that would be OK. But if other scientists do not agree to the oath, would you risk your country being at the mercy of a potential oppressor?

The other aspect of such an oath would be to protect against fraud and plagiarism. I realize there have been a number of frauds in science, but there are frauds everywhere. There are probably fewer frauds in science than the rest of society. People who commit scientific frauds are plain crazy; how could they think they would not be discovered? Other people will try to repeat the experiment and the fraud will be discovered — it is inescapable. ■



I don't agree with Laureate Lehn's answer. During the pursuit of science, we should pledge an oath and then follow it. Many scientific advancements have endangered human lives. I started as a medical doctor, and I took the Hippocratic oath to swear that I will not do anything that would harm my patients. Then I became a research scientist and I realised the potential for scientists to cause harm. A 'scientific oath' would prohibit scientists from indulging in research that could harm human life: bombs or bioterrorist agents for example.

Ehsan Ullah, Department of Morbid Anatomy & Histopathology, University of Health Sciences, Lahore-Pakistan, who posed the original question on lindau.nature.com.



Ferid Murad

Heart of science

Biochemist at the George Washington University in Washington, DC, he shared the 1998 Nobel Prize in Physiology or Medicine for the discovery that nitric oxide acts as a signalling molecule in the cardiovascular system, prompting blood vessels to relax. Murad was born in Whiting, Indiana in 1936. His American mother was only 17 years old when she eloped with his father, an Albanian immigrant. His parents ran a restaurant, where he and his two brothers worked. Murad used to memorize customers' orders and mentally tally their bills, which he believed trained his memory and maths skills.

How important is it for young doctors and medical researchers to think about the bigger questions: the essence of truth or the existence of God?

Scientists by nature have to be curious to answer questions of nature — to discover how things work. The beauty of science is that once you've answered a question, that leads to further questions, sometimes more important ones. Doctors are taught scientific enquiry in medical school, but it's not required that they be scientists. Yet if they are taught this well, it should improve their skills.

Science is about seeking the truth. The existence of god is irrelevant to a scientist, as is

his or her faith. It is possible to have faith and be a scientist at same time; it is also possible to be an atheist and a scientist at the same time.

There are some researchers, however, whose faith and religion tend to distort the facts. That's not going to lead to high-quality science. For example, some people's religion makes them reluctant to perform embryonic stem cell research. Many think it's unethical — and some politicians have made it illegal, but that's foolish. These are tissue samples that will otherwise be incinerated. They present an opportunity to do good biology and get information that is not otherwise available.

Science is all about getting to the facts — to

information: how creatures are 'created' and evolve, including on other planets too where there could be life. Extra-terrestrial life will be interesting to prove one way or another. There are so many thousands of planets in this galaxy that ours can't be the only one to develop life.

NATURE.COM
Nature Video of
Murad advising a
young researcher:
go.nature.com/a5eijf

F. CARTER SMITH/SYGMA/CORBIS

Incidence of diabetes is increasing worldwide. How do we minimize this problem?

When I was a trainee in the United States in the 1950–60s, the incidence of diabetes was about 2%, today it's 7–8% — and in some subsets or minorities it is higher. The Pima Indians in southern Arizona have an incidence of about 70%. It's incredible. They have been researched by the National Institutes of Health to help understand some of the reasons underlying diabetes. We have learned over the years that there are multiple causes: genetics, infections that injure the pancreas and its ability to produce insulin, diet, exercise and obesity. It's a complicated and growing problem.

Most problems with diabetes are cardiovascular. As the disease modifies proteins in blood vessels it leads to atherosclerosis, and, in turn, compromises blood flow to the heart, limbs and other tissues.

Do you think efforts to control diabetes could learn from the example of cardiovascular disease, which is better managed now?

It isn't fair to imply that cardiovascular disease is going away. Frequency of mortality with cardiovascular disease has improved: we are better at treating acute heart attacks and arrhythmias; we have better-trained paramedics and better-equipped emergency rooms. But people who have had heart attacks now live with injured heart muscle, which predisposes them to congestive heart failure. Plus they will have endothelial dysfunction of the blood vessels because they don't make enough nitric oxide. So there will be serious cardiovascular problems in the future.

We are better at controlling and treating hypertension, which is a big factor in cardiovascular disease. Incidence and frequency of cardiovascular disease is diminishing a little and cancer will soon overtake cardiovascular disease in frequency of mortality, but they are both still serious problems.

Life expectancy will continue to increase, although maybe not as rapidly as in the past 100 years following introduction of vaccines and antibiotics. However, life expectancy in the US is lower than in many other Western countries because of our style of fast living, fast foods, stress, etc.

Should medical science draw on insights from psychology, behavioural and social science to try to change detrimental human behaviours?

They can all influence behaviour, but they won't cure diabetes. That requires sophisticated medical research. The inheritance of diabetes is



Murad takes some time to pass on his experience to young researchers at the 2011 Lindau meeting.

probably not just a single gene but a concert of them. When there are multiple genes participating, it's often very difficult to sort out.

We were hoping that the human genome project would provide a lot of answers. Yet, it hasn't provided them all because there are multiple genes and factors that participate in these diseases.

Do you think antibiotic resistance is a big threat?

We're giving antibiotics to livestock. That is nonsense: it is creating resistant organisms, because the antibiotics are not being used to treat disease and the livestock owners are not dosing properly. Furthermore, we don't use these drugs in combination to eliminate organisms. We're always searching for better ones because we're not using them properly.

Viruses and bacteria are pretty clever — some more than others. Look at the effort and expense to develop treatment for HIV. We're using

multiple drugs to treat patients, but they still have latent virus hibernating somewhere. We're afraid that if we stop treatment the virus will come back. We can slow it down and make people live longer, but we haven't cured it yet.

Do you always think and behave scientifically?

I'm a workaholic. I love science. I think about it almost all the time. Even when I try to relax: watching TV or doing something else, I can only do that for 5 to 15 minutes, then I get distracted thinking about experiments. It has been disruptive to family life. I have five children and I probably haven't spent enough time with them over the years. When they were younger, I would always take 2 to 3 weeks in the summer to go camping with them, and I tried to be home every day for dinner. But even if I made it, I often went to my study or back to the lab afterwards.



This is an awesome answer, because a lot of scientists will only say that science is a passion and it should not matter if you spend every Saturday night in the laboratory. I don't think that's healthy for the mind. I completely appreciate his wish to spend more time with family and to face less intrusion.. Life should be about equilibrium between love, work, family, religion and your hobbies; they all make you happy.

William Omar Contreras Lopez, a Columbian neurosurgeon and PhD student in molecular neurosurgery in Freiburg, Germany, who posed the original question on lindau.nature.com

What did you learn from your mentor, and what do you think your students would say they have learnt from you?

I was fortunate because I had a long period of training and had many mentors who were excellent. They tended to give me a lot of freedom; they were there to help me and answer problems and review ideas. I try to recognize the strengths of each of them and come up with some hybrid that I can be as a mentor, to use the best features of each.

My first mentor in graduate school, Earl Sutherland Jr — who received a Nobel prize in 1971, taught me a lot about creativity. Research is not doing what's been done before — that's confirmation. Research is doing something that's never been done before — that's creativity.

Conversely, what didn't you learn from your mentors?

They all had a lot of scientific and personal strengths, however none of them knew anything about the drug development industry, business or finance. I had to learn a lot of that on my own. Some of the business folks I've met along the way have taught me a lot about businesses and what it takes to get something done. That's very different from working in a lab.

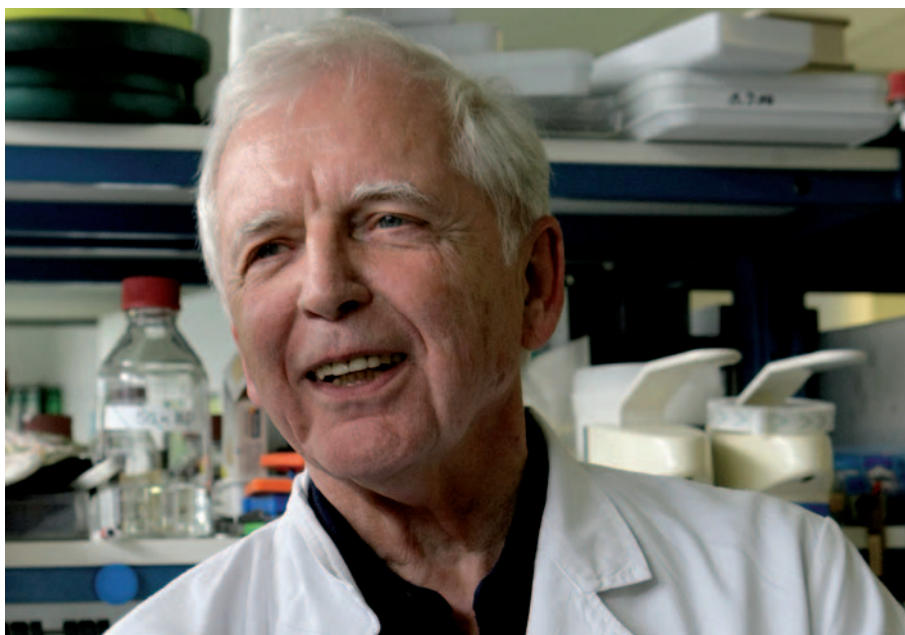
Is there a difference in the types of science that public versus private organisations can or should do?

I've run one company and helped friends and colleagues create about seven others. Academics do science because they love it. Of course, you have to be successful and get grants, be published, be recognized and get promoted. But basically you really enjoy it. In industry you also enjoy it, but you don't necessarily have to publish; companies value patents. And the rewards are to the team not the individual. Industrial science is much more of a team effort.

Academia and industry can learn from each other. Many projects and problems require collaboration between the two. I don't think an academic can find the funds to take a compound into clinical trials. But industry doesn't necessarily have the skills to find the target to start the process of drug discovery. We need more collaboration. The problem is that people are sceptical and tend not to trust each other.

➔ Is there a downside to winning the Nobel prize?

Yes. My wife would agree. It results in a lot of travel. Everybody expects you to know everything about everything. They don't realize that you really have a discrete specialization in one area; they think you can do anything — like advise presidents to solve problems in education. Also, you're on the internet, so everybody knows about you and you lose your private life. That disturbs me a bit. Everywhere you go there will always be someone who recognizes you as a Nobel laureate, and that can be hard to cope with. ■



Harald zur Hausen

Science addict

Virologist at German Cancer Research Centre in Heidelberg. Joint winner of 2008 Nobel Prize in Physiology or Medicine for discovery of the role of human papillomavirus (HPV) in causing cervical cancer. zur Hausen was born in 1936 in Gelsenkirchen-Buer in Germany, an area that was heavily bombed during the Second World War.

At Lindau you gave a talk suggesting that colon cancer might be linked to a virus in cows, which could survive in raw or rare beef. How realistic is the threat that we might be harbouring unknown cancer-causing viruses?

It is a big problem, and it's a needle in a haystack. With the advance of sequencing techniques, we can expect the discovery of a large number of novel viruses in different human tissues, among them some cancer causing agents. But some viral infections occur at very low frequencies, so it is hard to find them. I am convinced there are a large number still awaiting detection.

The gut is a huge repository of microbes. Many viruses in the gut are bacterial viruses — bacteriophages. So it is a complicated story that will take a long time to disentangle. Plus it is a long way from initial isolation of a virus to final identification and determination of whether a virus is pathogenic or oncogenic. We need functional tests to determine the relationship between a virus and cancer. For instance, we started to work experimentally on HPV in 1972. We isolated HPV strains 16 and 18 (the oncogenic ones) in 1983 and 1984, so that took 11–12 years before we could be sure that those were the causative agents.

Many of the world's most significant health challenges are caused, at least in part, by human behaviour. What is the role of science/medicine — along with psychology, behavioural and social sciences — in helping to change detrimental human behaviours?

All these sciences have to interact to some degree. If we could all avoid such things as smoking, overeating and heavy drinking then we have a good chance of reducing the burden of cancer globally. But, if we fully understand the factors underlying how smoking contributes to lung cancer, or how obesity can lead to diabetes or breast cancer, then this is the contribution from science and we have an opportunity to interfere through medication or in other ways.



If we consider that, for instance, smoking is a very bad idea, it then seems a little odd that we should seek to find cures to the many and complex diseases it causes at the expense of time, money and attention being dedicated to the behavioural sciences that might be more effective in reducing the incidence of the behaviours that lead to these diseases. There is a lot of effort developing evidence-based theory and techniques to change health-related and risky behaviours: education alone isn't sufficient.

Chris Martin, neuroscientist, Department of Oncology, University of Oxford, United Kingdom, who posed the original question on lindau.nature.com.

NATURE.COM
Other laureate
Q&As online:
go.nature.com/a5ejf

At first, pharmaceutical companies were dismissive of an HPV vaccine — has your relationship with them improved?

My relationship with the pharmaceutical industry is not optimal. Of course, they're mainly interested in selling their products, and they are reluctant to develop products without a market. And for certain products, if there is not governmental support for development, then very little will happen.

Fortunately, there are foundations that do good work along those lines. They help support treatments for some diseases that do not play a significant role in the West. Rotavirus, for example, has a huge impact on mortality in less developed countries. Because of the support of these foundations, vaccines for rotavirus are reasonably cheap and can be distributed on a larger scale.

What has your relationship been like with your mentors over the years?

I don't think I have had one particularly influential mentor. Each of my mentors gave me a lot of freedom, which I enjoyed very much, so I have some difficulties to pick one out. But, I had a good time with Werner and Gertrude Henle at the Children's Hospital of Philadelphia.

While at the University of Düsseldorf early in my career, I wasn't getting much advice from my supervisor. I would go to him with crazy ideas, and he would say "sounds interesting, why don't you try it out?" I hated it at the time because I felt like I wasn't being trained. In retrospect, however, it was a very good, creative period. I worked on a lot of nonsense, but I was able to make my own mistakes. In addition, I had the freedom to look into a lot of different areas.

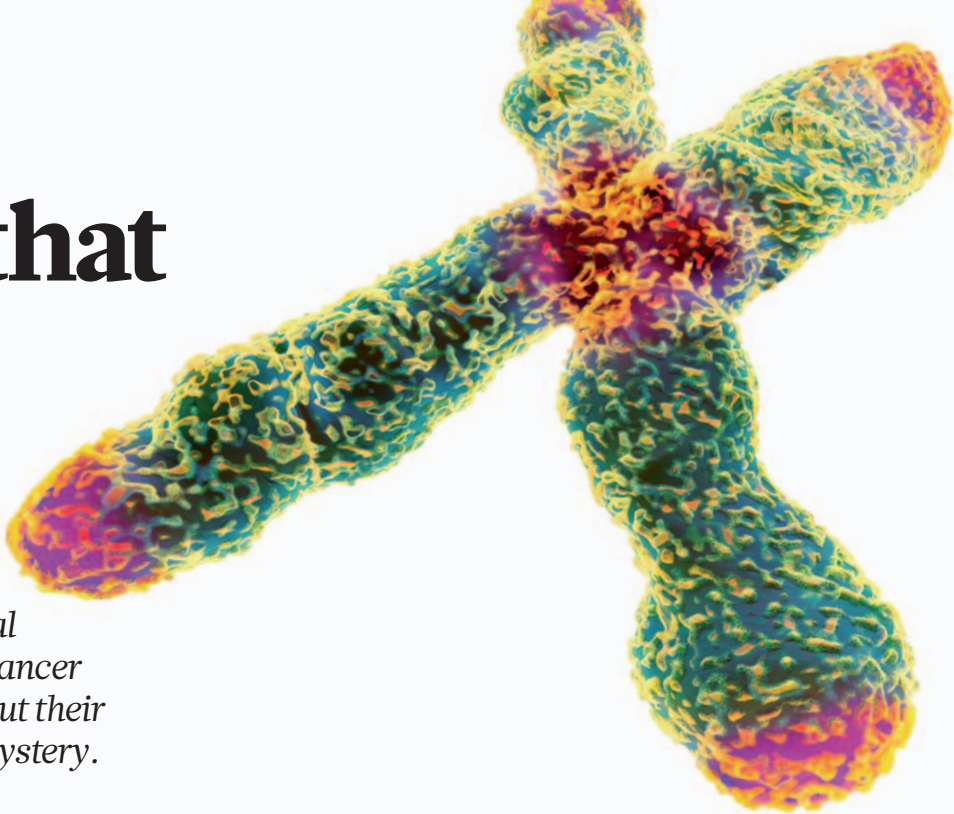
In turn, I tried to give my students a bit more freedom and to encourage them to develop their own ideas. These days I'm not accepting students any more as I'm 75 years old and I think it's a good time to stop taking them on.

Do you still spend long hours in the lab?

Whenever I can, including weekends. Science is like a drug — I'm an addict! ■

All's well that ends well

Elizabeth Blackburn gave the first lecture at the 2011 Lindau meeting, describing her Nobel prizewinning work on telomeres. These chromosomal caps are known to play a role in cancer and are implicated in ageing — but their full biological utility remains a mystery.



BY MICHAEL EISENSTEIN

TAGGG. TTAGGG. TTAGGG. And so on. And so on, hundreds and hundreds of times. It seems hard to believe that such mind-numbing repetition could conceal anything of interest.

Yet these nucleotide sequences, which cap the ends of every mammalian chromosome, appear to act as both an important bulwark against tumour growth and a buffer that delays onset of a number of the degenerative processes that take hold as we age. The Nobel Committee recognized the importance of these chromosomal end-pieces — called telomeres — when they awarded the 2009 Nobel Prize in Physiology or Medicine to Elizabeth Blackburn, Carol Greider and Jack Szostak for their trailblazing efforts to map details of telomere function and the mechanism by which these structures are generated and maintained.

When Blackburn began investigating these phenomena in single-cell organisms in the late 1970s with her mentor Joseph Gall, “it was much more of an esoteric, yeast and protozoan biology field,” says Jerry Shay, a cell biologist specializing in telomere research at the University of Texas Southwestern Medical Center in Dallas. “Now it’s really becoming a major area of investigation, and there are lots of new and exciting work that’s going to be coming out.”

The story that has emerged is not a simple one. For example, truncated telomeres can be seen as

“If I had short telomeres, I would probably be a more compliant patient.”

a harbinger of both good and bad: slowing cancer cell division but also impeding natural tissue repair processes. In both humans and mice, short telomeres are implicated in disease progression

and ageing, yet scientists are still struggling to understand how measures and manipulation of these enigmatic end-caps might benefit human health.

PLAYING A DOUBLE GAME

Owing to a quirk in DNA replication, chromosomes become slightly shorter each time they are copied. Excessively short telomeres could disrupt essential DNA sequences, and introduce the risk that cellular repair mechanisms might mistake frayed chromosome tips for broken DNA strands and graft these loose ends together. In the late 1930s, Barbara McClintock — a Nobel prizewinner in 1983 — showed that unprotected chromosome ends are prone to cycles of fusion and breakage, resulting in severe disruption and rearrangement of chromosomal structure. The main function of telomeres is to prevent such problems: these repetitive caps of DNA erode over many rounds of cell division while insulating essential sequences against injury.

Telomeres are made and maintained by the enzyme telomerase. Telomerase is intermittently active within populations of adult stem cells, which contribute to the maintenance and repair of various tissues throughout the body. Most other adult cells do not produce detectable telomerase, so their telomeres become shorter with every division. When telomeres reach a critical length, they trigger crisis management mechanisms such as p53, a tumour-suppressor protein that monitors DNA integrity and can induce cells to either self-destruct or enter a dormant state known as senescence (see ‘A tale of two ends’).

Because chromosomal damage can promote uncontrolled cell growth, one might expect that rampant telomere erosion could introduce the same risks. Fortunately, p53 halts this process by preventing cell division. Greider and others have shown that senescence induced by telomere

shortening might offer protection against tumour formation. “It’s clear that, in general, short telomeres can limit cell division and stop the growth of tumours in several mouse models,” says Greider, a molecular biologist at Johns Hopkins Medical School in Baltimore, Maryland. “The question is to what degree that translates to different human settings.”

If genetic mutation disrupts p53 function, as occurs in a variety of cancers, a key safeguard is lost and short telomeres become a major liability. Cancer biologist Ronald DePinho studied shortened telomeres in the absence of p53 activity in mouse models. “We saw an increase in cancer,” says DePinho, president of the MD Anderson Cancer Center in Houston, Texas. What’s more, although cancer-prone mice with normal length telomeres usually develop different cancers than are typically observed in older humans, DePinho’s short-telomere mice developed tumours in the breast, colon and skin. “These mice changed their tumour spectrum towards a humanized spectrum of epithelial cancers,” a finding that he says highlights the importance of telomere length in human cancer.

Tumour progression is limited without telomerase. “The almost universal way in which cells continue to divide and become an advanced, robust malignancy is by somehow upregulating telomerase, so they can maintain the telomeres that are being lost with every division,” says Shay. Indeed, Shay and long-time collaborator Woodring Wright, also at Southwestern Medical Center, have found that telomerase might be active in 90% or more of human tumours. Researchers once thought that this restored activity required full reactivation of the dormant telomerase gene in the tumour cell. More recent findings suggest that tumour growth is instead initiated in stem cells that already produce telomerase, necessitating only a small bump in

expression rather than a dramatic upswing. “The vast majority of tumour cells have very short telomeres,” says Wright. “They maintain length, and are not actively elongating.”

Given that cancer cells tend to exhibit steadier overall telomerase activity than healthy stem cells, this enzyme could offer a useful target for a variety of cancer treatments. Menlo Park, California-based biotech company Geron is in the midst of several phase II trials to test the efficacy of imetelstat, a telomerase inhibitor, against lung cancer, breast cancer and myeloma. According to Stephen Kelsey, Geron's chief medical officer, this molecule is most effective against the rapidly dividing, immature tumour cells that some cancer researchers have tentatively labelled ‘cancer stem cells’. He suggests that drugs like imetelstat might work best following surgery or chemotherapy, when surviving tumour cells are proliferating aggressively in an attempt to rebuild.

Inhibiting telomerase is expected to halt tumour growth either by forcing the cells to enter senescence or through excessive genomic damage, although this has not been demonstrated in humans. Geron hopes to have preliminary efficacy data by late 2012; initial safety studies at clinically useful dosages have found no evidence that the drug has lasting adverse effects on healthy telomerase-expressing cells. “Generally, tumour cells have shorter telomeres, which would provide a therapeutic window where they get selectively killed off,” says Calvin Harley, former chief scientific officer of Geron and now CEO of Telome Health, a diagnostics company also based in Menlo Park.

THE LIMITS OF REGENERATION

Aside from the cancer connection, some of the greatest excitement in this field focuses on how telomere shortening contributes to human age-related disease. “I wouldn't say that cancer has moved to the back seat, but rather that the connection to human age-related disease has moved to the forefront in the last few years,” says Mary Armanios, an oncologist at Johns Hopkins.

An extremely rare disease known as dyskeratosis congenita (DC) was the first condition to be directly linked to mutations in genes involved in telomere maintenance. Patients with DC lose the capacity to produce new blood cells owing to a collapse of their precursor stem cell population, a state known as bone marrow failure. More recently, telomerase gene mutations were also linked to risk of idiopathic pulmonary fibrosis (IPF), a more common, fatal disease characterized by extensive scarring in the lung.

Both DC and IPF appear to represent a failure by the body to repair aged or damaged tissues. “We know that telomerase is maintaining telomeres to enable self-renewal in stem cells, and when that goes awry it leads to disease,” explains Steven Artandi, a haematologist at Stanford University in California. “In many cases these patients appear normal at birth; then, as they enter late childhood or early adulthood, depending on



Laureate Elizabeth Blackburn, a pioneer of telomere research, chats with a young researcher in Lindau.

the mutation, they start to experience failure of different tissue compartments.”

People who inherit these telomerase mutations might also inherit prematurely shortened telomeres, leading to earlier and earlier onset of IPF. Studies by Armanios and colleagues have demonstrated that this intergenerational shortening is associated with increasingly severe pathology. “Older generations develop IPF, but as the disease gets worse with each generation, the younger generations develop a bone marrow failure disorder called aplastic anaemia,” she says.

Telomere shortening appears at least partially culpable in a growing list of diseases. Early attempts to characterize the physiological effects of telomere shortening hit a wall when researchers found that mice lacking telomerase appeared to suffer no ill effects. Closer inspection revealed that standard inbred strains of laboratory mice are born with telomeres several times longer than in humans, limiting the effects of age-associated shortening. Scientists have overcome this limitation via serial breeding of telomerase-deficient animals and use of newly derived mouse strains with human-like telomere lengths. With these short-telomere strains, it is now possible to model in mice many diseases previously intractable to *in-vivo* investigation. For example, standard laboratory mice “don't develop liver cirrhosis

“Then we flipped the switch back on, and the animals had improved cognition, they became fertile and so forth.”

even after you ablate the liver many times,” says DePinho. “Once we generated mice with shorter telomeres, these mice developed florid liver cirrhosis.” This result, DePinho explains, showed that shortened telomeres are “integral

to the pathogenesis of this disease state.” Similarly, Artandi and Stanford colleague Helen Blau showed that developing mouse models of Duchenne muscular dystrophy requires mice with short telomeres². “If you take telomerase away, short telomeres impair stem cell function and you suddenly reveal the full effect of the Duchenne mutation on muscle physiology in mice — and it looks very much like it does in humans,” says Artandi.

COUNTING THE DAYS

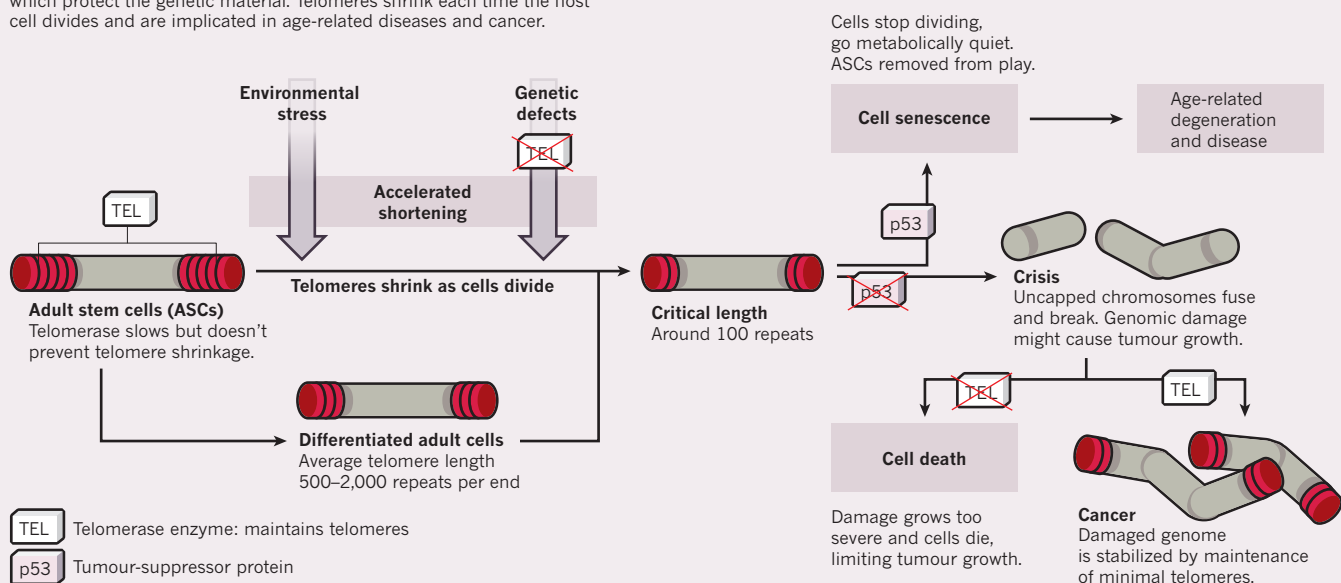
Many of the above pathologies seem to recapitulate the tissue degeneration we experience as we grow old, and the idea that telomere shortening helps drive the ageing process has been a central focus of the field.

Studies in mice, at least, have suggested that telomere maintenance can have an impact on both quantity and quality of life. “Work in mouse models clearly indicates that having shorter telomeres than normal is a cause of ageing,” says Maria Blasco, a cell biologist at the Spanish National Cancer Research Centre (NCIO) in Madrid. Using genetically manipulated animals, Blasco and colleagues have shown that increasing telomerase at the same time as increasing expression of the p53 pathway could “extend the lifespan of mice by 40%” (ref. 3). More recently, DePinho's team assessed the impact of reactivating telomerase in mice whose telomeres had reached a critical length⁴. “The mice were not in good shape — equivalent to somebody in their 80s or 90s,” he says. “Then we flipped the switch back on, and the animals had improved cognition, they became fertile and so forth; this told us that there's a point of return for tissues, even those in a severe state of degeneration.”

Indeed, the correlation of ever-shortening telomeres with the passing of the years lends

A TALE OF TWO ENDS

Chromosomes are capped by telomeres, a series of TTAGGG repeats, which protect the genetic material. Telomeres shrink each time the host cell divides and are implicated in age-related diseases and cancer.



itself to a tidy narrative in which telomere length somehow records biological age, like rings in a tree stump. However, scientists working with telomeres caution that this is a gross oversimplification. “There is clear evidence against the concept that telomeres are the sole determining factor in ageing,” says Greider. “There may be five, six or seven different pathways that play into what we consider as ageing, and telomeres play one role in that process.”

All the same, the tantalizing association between short telomeres and various diseases has persuaded several leading researchers to launch companies that measure telomeric length in blood cells for prognostic and diagnostic purposes. Repeat Diagnostics, based in Vancouver, Canada, was founded in 2005 by oncologist Peter Lansdorp to help clinicians identify patients with genetic mutations affecting telomere function. Two others companies are reaching out directly to consumers and healthcare providers: Blackburn and Harley teamed up to launch Telome Health in March 2011; Blasco and colleagues founded Life Length in Madrid in December 2010. Indeed, Life Length claims to be able to measure ‘biological age’ and recommends annual measurements of telomere length.

The customer focus of these latter companies has been the subject of debate. “People want to know their telomere length because there are more and more papers being published that show that the lowest percentiles of telomere length have a higher risk for disease,” says Blasco. Harley points out that even if the association between telomere length and specific conditions is ambiguous, people might still be able to make lifestyle decisions based on these measurements. “It allows you to monitor your health more closely and work with your doctor to get more specific

information about where it is that you’re most likely at greatest risk,” says Harley. “If I had short telomeres, I would probably be a more compliant patient.”

Other scientists are concerned that too little is known about the impact of short telomeres for such a measurement to be meaningful. “It could certainly be useful — I just think it’s premature,” says Greider.

MEASURE FOR MEASURE

Clarifying the relationship between telomere length and disease will require more prospective studies of large cohorts of people. “Unfortunately, there’s a lot of data out there that I think is very questionable,” says Lansdorp. “The studies that have been done are primarily epidemiological.” As a consequence, he says, “a lot of these observations may be correlative and indirect”.

Methodology is also a point of contention. Telome Health, in common with most epidemiological studies performed to date, employs a DNA amplification-based technique that quickly determines average telomere length within cells, but which might be prone to technical artefacts. Blasco, Lansdorp and others use more time-consuming and costly techniques that use fluorescent probes to examine individual telomeres within a cell, uncovering individual outliers that might be masked by averaging. “Each cell has 92 telomeres, and each telomere has its own unique length,” says Armanios. “The shortest telomere is the one that determines the phenotype.” These measurements are typically taken from white blood cells, which are easy to obtain and closely reflect the telomere length of the stem cell reservoir from which they emerge; however, it remains unclear whether these measurements reflect telomere length throughout the body.

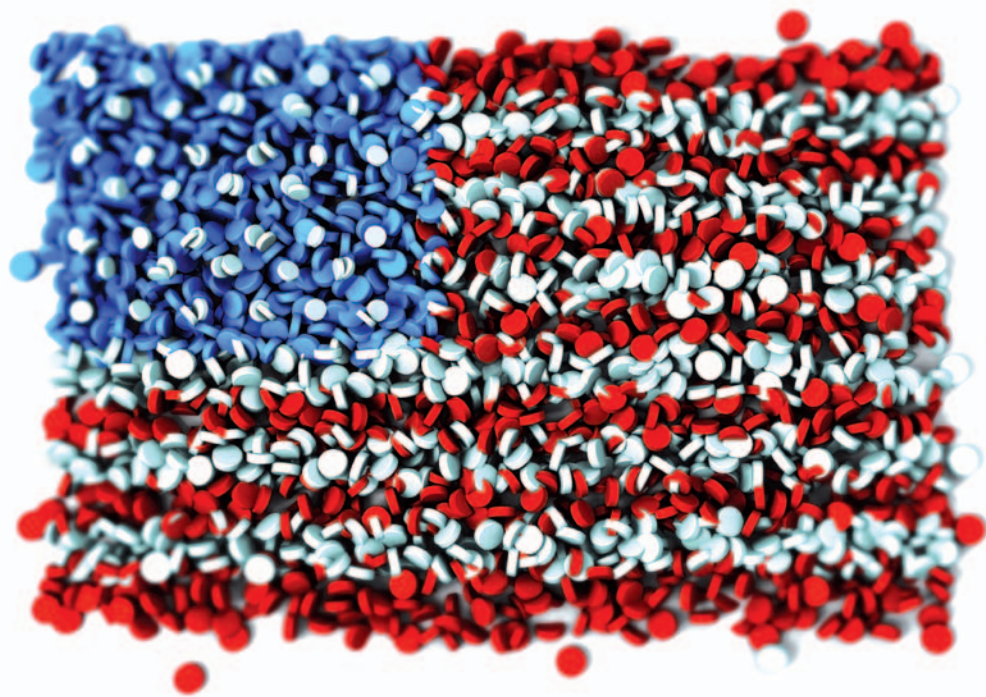
Telomere length can be affected by environmental factors, further complicating the interpretation of length measurements taken at any one time. Studies led by Blackburn and Elissa Epel at the University of California, San Francisco, have uncovered evidence that telomeres undergo accelerated shortening in direct response to lifestyle and emotional stress, such as from post-traumatic stress disorder or major clinical depression⁵. “When this first came out I was sceptical, but I’m starting to believe a little bit of this stuff now,” says Shay. “I’m still on the fence about how much it really means, but I believe it shows that the shortest telomeres can be modified by environmental stressors.”

Shay, who is also a scientific advisor for Life Length, hopes that the data collected by the direct-to-consumer companies will shed light on the influence of lifestyle and other mysteries by providing population-wide data from long-term, longitudinal telomere profiling studies that might otherwise be too expensive to support via public funding.

From DePinho’s perspective, efforts to figure out the role of telomeres in cancer and ageing are a race against the clock. “By 2025, there will be 1.2 billion individuals over the age of 60,” he says. “We need to understand the circuitry of ageing, and we need to understand how to control it so that we can increase the years of healthy living, diminish age-associated diseases and maintain productivity in our aged population.” ■

Michael Eisenstein is a freelance journalist based in Philadelphia, Pennsylvania.

1. Parry, E.M. *et al.* *Blood* **117**, 5607–5611 (2011).
2. Sacco, A. *et al.* *Cell* **143**, 1059–1071 (2010).
3. Tomás-Loba, A. *et al.* *Cell* **135**, 609–622 (2008).
4. Jaskelioff, M. *et al.* *Nature* **469**, 102–106 (2011).
5. Wolkowitz, O.M. *et al.* *PLoS One* **6**, e17837 (2011).



TRANSLATIONAL RESEARCH

The American way

The United States publishes more biomedical research papers than ever before, yet drug development is stagnating. Several new initiatives aim to turn this knowledge into new remedies.

BY AMY MAXMEN

On being asked to grade the United States on its performance in biomedical research, Bill Chin, executive dean for research at Harvard Medical School in Boston, Massachusetts, responded: “If the measure describes how much we understand about disease, I think we’re on a good road. If it’s how often we turn basic science ideas into potential medicines, we aren’t doing that well.” For Chin and other scientists, discoveries don’t matter tremendously if their potential goes unrealized. Although research papers by US scientists document great strides in understanding human physiology, genetics and disease, a failure to efficiently translate those findings into diagnostic tests and medicines foment frustration among researchers and the public alike.

Advances in basic science abound. Of the 196 Nobel laureates in physiology or medicine since 1901, 46% have been US researchers. According to the National Institutes of Health (NIH) in Bethesda, Maryland, sequencing the first human genome took five years of work and US\$450 million. Today it can be done in 90 days and for US\$9,500. By the end of 2011, it is estimated that

North American scientists will have sequenced 9,000 genomes — an indispensable resource in genome-wide association studies that attempt to link genetic variants to common maladies. Although knowledge of these variants has yet to make a substantial impact on patient care, the trickle into the clinic has begun. For example, hepatitis C patients can now be tested for a variant of the *IL28B* gene, discovered by researchers at Duke University in Durham, North Carolina, to determine whether they will respond favourably to the gruelling treatment.

“Anyone who says there hasn’t been progress hasn’t been paying attention,” says Mary Woolley, president of Research!America, an advocacy group based in Alexandria, Virginia, committed to making biomedical research a higher national priority. “We’ve made progress in women’s health,

including getting women into clinical trials,” she says. “We’ve made progress in medicine,” Woolley adds, citing cancer and heart disease in particular. Between 2001 and 2007, the annual mortality rate from cancer dropped by 9.1% and by 23.5% for cardiovascular disease (see ‘Getting on top of cancer’). However, these positive trends are failing to convince the public. “Our polls show that two people to every one believe we aren’t making enough progress in biomedical research, and that’s kind of disturbing,” says Woolley.

This lack of belief might reflect frustration at the poor rate of return. According to Research!America, US national expenditure on health reached US\$2.6 trillion in 2010, representing 17.8% of gross domestic product (GDP) — a higher proportion than any other country. Yet the United States ranks middle-to-poor in measures of health outcomes, including longevity and infant mortality (see ‘Life expectancy’ and ‘Infant mortality’ graphs). And although several thousand reports claim particular genes or proteins have disease-fighting potential, the US Food and Drug Administration (FDA) has approved just 111 novel drugs in the past five years, slightly fewer than the five years before that (see ‘Drug approvals stagnating’). Likewise, the pace and price of drug development haven’t improved despite technological advances. Crucial molecules discovered in disease pathways have a 2% chance of being translated into a therapy. Any drug that makes it through the pipeline typically takes 13 years and more than US\$1 billion, including the cost of the failures. This is an inefficient process, to say the least, says Francis Collins, director of the NIH, who also addressed the 2011 Lindau meeting. “An engineer looking at this would go, ‘What? You’re going to base the future of human health on a pipeline that works like that?’”

Despite advances in understanding the pathogenesis of disease, Collins says the translation of these findings into clinical applications remains a slow, expensive and failure-prone endeavour. The blockages in the biomedical pipeline need to be identified and dissolved. And, given that the economy is weak, the US federal budget is in distress, and 10 blockbuster drugs are about to come off-patent to the tune of more than US\$50 billion in lost sales, drug developers face an especially challenging time. “The science has never been more exciting with the potential to revolutionize human health,” says Collins. “But support for this science has never been as threatened as it is right now.” Consequently, biomedical leaders in the United States recently launched a bevy of new programmes to accelerate drug discovery and attempt to reverse the decline in the pharmaceutical industry.

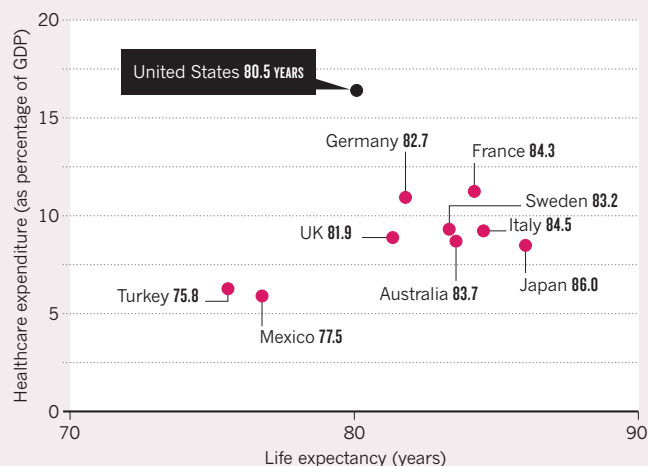
At about US\$45.9 billion per year, the federal government supports one-third of US biomedical research. Most of these funds are channelled into the NIH. Pharmaceutical and biotech companies fund most of the rest, spending about US\$76.5 billion in 2010. Universities,

LINDAU 2011

Each year the Lindau Meeting of Nobel Laureates in Germany has an international sponsor and a scientific focus related to a Nobel Prize. The United States and physiology or medicine took centre stage in 2011. *Nature Outlook* explores the intersection of these areas in a special report.

LIFE EXPECTANCY

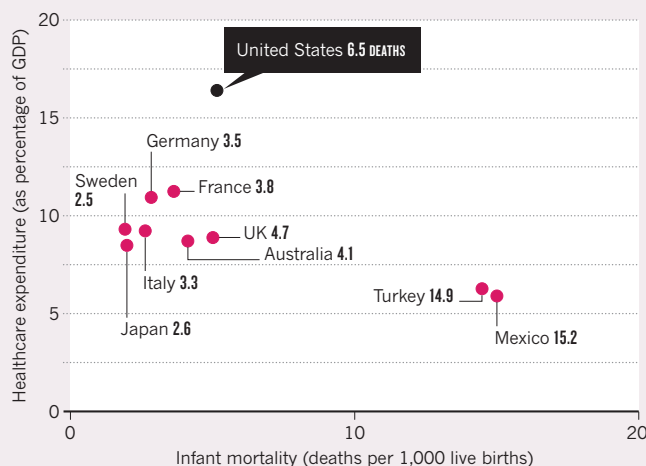
As a percentage of GDP, the United States spends considerably more than any other country on healthcare, yet this does not result in the highest life expectancy.



Data from 2008

INFANT MORTALITY

Many countries with lower healthcare spending have a lower rate of infant mortality than the United States.



philanthropic foundations and other sources also sponsor some research and development (R&D). Altogether, R&D accounts for less than 5.5% of the nation's health expenditure; hospital care and clinical services account for a much bigger slice of the pie.

To sustain their R&D expenditure, pharmaceutical firms need new drugs to sell. "We really are not the bottomless pit of money we were when I entered this business 25 years ago," says Charles Lunn of Merck Research Laboratories in Kenilworth, New Jersey. "If a programme is unsuccessful, we lose a lot of money. So when we go out looking for new opportunities to collaborate with the academic community, we are very careful."

RISKY BUSINESS

Kyle Palmer, director of research at Redpoint Bio, a small biotech company in Philadelphia, Pennsylvania, experienced this risk adversity firsthand as his company nearly folded. In the late 1990s, neuroscientist Robert Margolske founded the company, hoping to bring taste-modulating compounds to market. A few years later, researchers discovered that some of the compounds modulated insulin, and therefore held potential as new therapies for type 2 diabetes. Redpoint Bio was awarded a patent that granted exclusive rights over several compounds involved in taste- and insulin-modulating pathways, and began testing the compounds in mice. Then everything changed. "Just as I thought we were poised to make all kinds of discoveries, with several big drug companies interested, the economy crumbled and the rug was pulled out from under our feet," says Palmer. With no promise of investment, Redpoint Bio shelved its compounds. "It seems like pharma are dropping out of basic research," says Palmer. "They're farming out some preclinical research to biotech and academic operations, so biotech

has responded by changing model from early phase drug discovery to providing services to pharma like *in-vivo* testing, and that's exactly what we've done. We've auctioned off our drug discovery arm."

Lunn explains that pharma now need a mountain of evidence before investing in turning a discovery into a therapy. This is partly because biology has turned out to be more complicated than previously thought. "You can make up a good story about how any number of targets might have a useful benefit, but human biology is so complex that proving their validity requires a lot more information than is asked for by premier journals," he says.

The barren stretch between the abundant fields of molecular discovery and the sparse world of late-stage clinical trials has been dubbed the 'valley of death'. Traversing it means that academic researchers and industrial drug

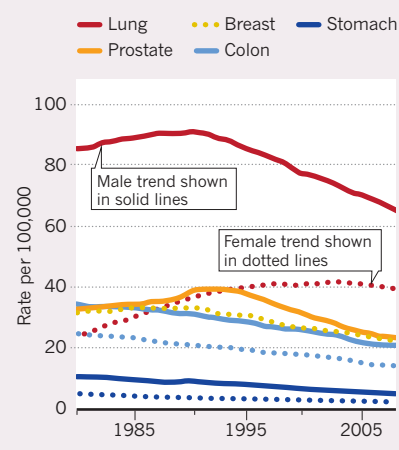
developers must move their domains closer together. This will require changes in incentives. For instance, academic researchers are rewarded for reporting novel discoveries rather than for replicating experiments in different animal models to support their initial findings. Eric Perakslis, at Johnson & Johnson Pharmaceutical Research & Development in Raritan, New Jersey, says that current practice leaves too much risk for pharma. Hundreds of posters at conferences claim that a gene or protein represents an effective antitumour target, he complains, without adequately showing what type of tumour it's most active in. Thus, he says, "you're talking about a few people in pharma making a billion dollar decision based on low correlation data".

On the other side of the valley, pharma impede translation when they file early for intellectual property (IP) that grants them exclusive rights to work on a compound before they have proof that it works in clinical trials, says Chas Bountra, a translational medicine scientist at the University of Oxford in the United Kingdom. If a company with IP puts a project on hold (as Redpoint Bio has), its compounds become inaccessible. "We're taking a process that is already incredibly difficult," says Bountra, "and making it more difficult and more expensive."

Rules and regulations provide another point of contention. Bill Crowley, professor of medicine at Harvard Medical School, says the conflict-of-interest policies at many research institutions give off a message of industry-distrust to young investigators and thereby inhibit collaboration. Crowley explains that investigators with potential conflicts of interest are often lumped together even when their conflicts vary dramatically: a researcher who once shared data with a company must complete the same lengthy paperwork as an investigator who is paid for ongoing, long-term

GETTING ON TOP OF CANCER

Over the past 30 years, the rate of cancer mortality in the United States has almost universally fallen.



consultation. Furthermore, “how much institutions ‘lump’ varies widely and in a most confusing way that makes life more complex for clinical investigators,” he adds.

Crowley also takes issue with tighter federal regulations on human studies, which he says have bogged down drug development without making studies much safer. “In the 1980s, consent forms for patients entering clinical trials were about a page long and almost invariably engaged a detailed discussion between the patient and me,” he says. “We now have a standard 19-page consent form written by lawyers, for lawyers. Most patients don’t bother to read it, and if they do, they don’t understand it.”

TIPPING POINT

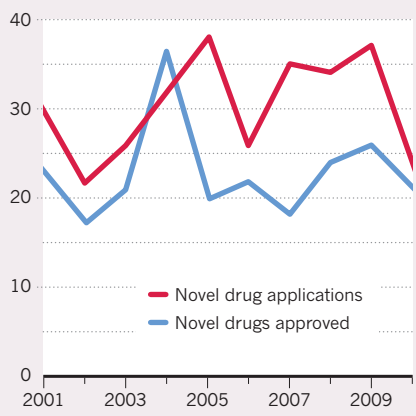
There have been many attempts to close the gaps in translational research over the years, with little success. “People are so frustrated on both sides of the divide that they are coming together and learning about each other’s incentives and disincentives,” says Johnson & Johnson’s Perakslis. “If my academic colleagues can’t get their grants renewed, that’s not good for me because if they can’t survive, they can’t be my partners.”

A number of partnerships between the public and private sectors have recently cropped up in various disciplines to free the flow through the biomedical pipeline. For example, the Foundation for the NIH, a nonprofit organization in Bethesda, Maryland that aims to accelerate NIH research, established the Biomarker Consortium in 2006. With participants from the NIH, FDA, industry and universities, and about US\$35 million in funding from nonprofit groups and pharmaceutical companies, the consortium has launched 12 projects to help take biomarkers through the large-scale clinical trials required to gain FDA approval. Combining their results allows all the partners to access cumulative data on thousands of patients.

Another such movement, the Structural Genomics Consortium led in part by Bountra, involves sharing information on molecular structures. Drug developers can increase the likelihood of finding an effective drug when they know the 3-dimensional structure of the human protein it’s intended to inhibit. Teams from academic centres and industry share data on a large number of drug targets, as well as reagents and assays. If no IP is filed until compounds complete early stage clinical trials, participating companies gain knowledge without hefty initial investments. “The beauty of these public-private partnerships, is that we’re saying let’s pool our resources and not worry about IP, and not worry about the slim chance this idea will generate money because frankly, most of the time it won’t,” says Bountra. He bets that prepublication sharing shouldn’t hurt academics either. Bountra has already posted on his website the structural details of a high-impact human membrane protein recently revealed

DRUG APPROVALS STAGNATING

Despite years of attention and investment, neither the rate of FDA approvals nor new drug applications have improved.



by his team, and a paper is in preparation. He’s confident that *Nature* or *Science* will not reject the paper on account of his openness. Indeed, both journals state that they allow pre-submission publication of information on nonprofit preprint servers, provided no conclusions are drawn. “I’m taking this risk because I’m less concerned about publications than I am about generating new drugs,” adds Bountra.

Scientists lacking Bountra’s optimism might take comfort in hearing that leaders at some top institutions including Harvard Medical School, Princeton University in New Jersey, and University of California, San Francisco, now look beyond the number of manuscripts an investigator has published to other factors, like contribution to collaborative research. And to ease fears about perceived conflicts of interest, Harvard Medical School revised its conflict-of-interest policies to more clearly delineate acceptable and prohibited activities. The revised guidelines encourage collaboration, explains Chin, by explicitly outlining allowable activities. “Faculty may continue to: conduct research sponsored and supported by industry; collaborate with industry

on research and serve as co-authors in these efforts; consult for industry; start biotechnology companies; serve on scientific advisory boards; and hold equity in most companies,” he states.

Repairing the pipeline is important to Collins; for the first time in the NIH’s history he has proposed closing one of its centres and reshuffling others to make space for the National Center for Advancing Translational Sciences (NCATS). NIH Investigators and scientists supported by NCATS grants will concentrate on the early stage drug development research that is often ignored by industry, including the development of methods and technologies to hasten target validation. Although these studies might not result in top-tier publications, they’re needed to bridge gaps in drug discovery. As for persuading scientists to focus on optimizing methods rather than discovering new targets, Collins believes being part of a large-scale effort is encouragement enough. When Collins led the Human Genome Project, he says peer-pressure proved a driving force. “You didn’t want to be the part of the team that missed the deadline. Everybody else was depending on your success for their success.” This feeling spread beyond the individual scientists. “There was a gradual recognition in academic centres that an investigator’s involvement with a successful scientific project was as important as how many papers that person had published,” he adds.

Over the next decade, success in closing this translational gap will largely be measured in tool development. Lee Nadler, dean for clinical and translational research at Harvard Medical School, says that helping clinicians prognosticate disease or adverse outcomes to medicine will be crucial. “If biomarkers told us who to treat, who not to treat, and who will react poorly to a drug,” says Nadler, “that would revolutionize drug development.”

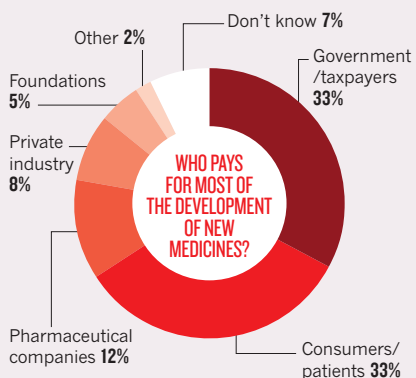
Better medicines will bring society closer to the ultimate goal of alleviating disease. However, they aren’t enough alone. Disease prevention is another factor; for this, federal support is essential. “If you’re a pharmaceutical company and your goal is to sell therapeutics, prevention is not your best idea of a business plan,” explains Collins. “But the NIH — with its mission to both understand the basics of how life works and apply that to the betterment of human health — has to have prevention front and centre.”

The public’s confidence in biomedical research is influenced by the health of the nation. Thus beyond lab work, the public health sector needs to work with health insurers, drug companies and the FDA to reduce the high cost of medicines and healthcare in the United States, says Harvard’s Crowley. Until all parties involved in health unite towards a common goal, no one scores an ‘A’ on the biomedical report card. Crowley says gravely, “The whole therapeutic programme in the US will go under if the jump from bench to bedside doesn’t go smoothly.” ■

Amy Maxmen is a freelance writer based in New York City

PUBLIC MISPERCEPTIONS

Most people in the United States do not know that pharmaceutical companies fund the majority of new drug research.





Ciechanover and young researchers listen to Geeta Negi present her work on diabetes neuropathy.

MASTERCLASS

The gathering brainstorm

In Lindau, a colloquy between a Nobel laureate and three students encouraged the young researchers to grapple with some of the biggest challenges in drug development.

BY KAT MCGOWAN

It was the stuff of dreams for a young researcher — a chance to be coached by a science superstar. The ‘masterclass’ that took place on the second day of the 2011 Lindau Meeting of Nobel Laureates was an experiment in which student science took centre stage. Three young investigators gave poster-talk style presentations of their current work to a Nobel prizewinner. The master, in this case biochemist Aaron Ciechanover (see page S4), critiqued each talk. The tone was set early: “Unlike many lectures here that ask about career development,” he proclaimed, “this is going to be about science.”

Ciechanover, a biochemist at Technion, the Israeli Institute of Technology in Haifa, shared the 2004 Nobel Prize in Chemistry for the discovery of the ubiquitin-proteasome system, the tightly regulated process by which cells tag and dismantle unwanted proteins. This protein quality-control mechanism is crucial to such basic biological phenomena as cell division and DNA repair, and likely to play an essential role in diseases ranging from Alzheimer’s, in which proteins aggregate, to cystic fibrosis, in which they are dismantled too quickly.

The three students he chose for the session all explored an aspect of protein folding or processing. But the session soon evolved into a

fast-paced lesson on the importance of thinking about how to stop diseases, rather than merely understanding them. Ciechanover’s take-home message: responding to today’s global epidemics requires the ability to convert scientific ideas into treatments that work.

SURFACING PROBLEMS

The young researchers’ talks focused on three major global health issues: Parkinson’s disease, diabetes and tuberculosis (TB). Ciechanover contended that they are all modern diseases. Parkinson’s disease, a neurological disorder caused by the loss of dopamine-producing neurons, is rare in people younger than 50 years of age. Until the past 100 years, few people lived long enough to develop it. The rapid global ascent of diabetes can be attributed, at least in part, to affluence: eating foods high in fat and sugar, with a sedentary lifestyle. And the trajectory of TB reveals both the power and the weakness of drug development: beaten into submission by powerful antibiotics in the mid-twentieth century, this infectious disease came roaring back in a menacing new drug-resistant form at the end of the millennium — a nasty reminder that medical triumphs can be short-lived.

First to present was doctoral student Lena Burbulla of the Hertie Institute for Clinical Brain Research in Tübingen, Germany. Burbulla

described how the quality control of mortalin, a mitochondrial chaperone protein, goes awry in Parkinson’s disease. Normally, explained Burbulla, mortalin assists in refolding misfolded proteins and protecting neurons from protein aggregation, oxidative stress and mitochondrial damage. Three variants of the mortalin gene have been found in Parkinson’s patients. Neurons in the Parkinson’s brain suffer from both mitochondrial dysfunction and protein accumulation, but the causal relationship is not clear; probing the effects of these gene variants could clarify the pathogenesis of the disease, said Burbulla.

Next up was Geeta Negi, a student at India’s National Institute of Pharmaceutical Education and Research in Punjab, whose research concerns the role of the transcription factor NF- κ B in diabetic neuropathy — a crippling complication of diabetes that leads to sensory nerve damage. There are no treatments that block the inflammatory cascade underlying diabetic neuropathy, so interfering with NF- κ B could be a promising approach, said Negi.

In the final presentation, Avishek Anant of India’s National Institute of Immunology in New Delhi, talked about drug targets in *Mycobacterium tuberculosis*, the TB pathogen. To better control TB, Anant said, clinicians need many innovations: new drugs to treat multi-drug-resistant TB, point-of-care diagnostic tests and a post-exposure vaccine. These needs can seem overwhelming, but Ciechanover reckoned this scope of issues was entirely appropriate for his masterclass: “We are not solving problems here. We are *surfacing* problems.” These three diseases, he said, are a problem now — and epidemics of the future: all are on the rise and none has an obvious cure.

Rather than focus on the specifics of the presentations, Ciechanover prompted a conversation about how best to design selective therapies. Using Negi’s NF- κ B idea as an example, Ciechanover identified a difficulty in targeting this protein complex: the promiscuous transcription factor is involved in multiple systems, including neuronal survival, the production of myelin (the protective sheath that encases neurons) and in the expression of genes that block cell death (or apoptosis). “This is one of the most complex transcription factors described,” he said. Turning to Negi, he asked: “How would you protect and spare its activity in one cell, and tamper with it in another?”

Ciechanover then turned Negi’s problem into a challenge for the audience, asking everyone to imagine themselves in the role of a researcher at a drug company. “I ask you to come up with solutions,” he urged, pointing at the audience; hands shot up all around the room. One young cancer researcher suggested attaching a toxin to an antibody that would target a specific antigen in the tissue. “Fantastic,” responded Ciechanover. Another proposal: targeting highly activated NF- κ B with an endogenous regulator that could generally down-regulate background levels of the

transcription factor. “It lacks specificity,” observed Ciechanover. Perhaps, elaborated the student, a vector that directed the regulator to the appropriate tissue, capable of being activated or shut down, would provide the necessary precision.

Ciechanover stalked around the stage, interrupting students to praise or scrutinize each idea in high-energy exchanges that enthralled the room. Students eagerly proffered suggestions, and there were many more raised hands than there was time to hear from each person. Ciechanover steered the discussion beyond the specifics of NF- κ B to tackle the broader problem of selective inhibition of biological processes. This is not just a problem of one transcription factor, he said, “this is a problem of medicine.” Cancer is the most obvious example: chemotherapy not only kills tumour cells, but does considerable collateral damage to other cells. How can this be avoided?

In the session that followed, Ciechanover offered criticisms and suggestions to hone each participant’s idea. One proposal was to deliver, via a viral vector, a protein that inactivates the target. The protein’s sequence would be encoded in the vector and controlled by a promoter — a stretch of regulatory DNA — that was tissue-specific, such that the protein would be activated only in the target organs. Ciechanover considered it, pointing out that promoters tend to get shut down by methylation, and therefore cannot act long enough to be effective. Another suggestion was to deliver an agent orally, but turn it into a working drug only at the target site with a technique like photoactivation. Ciechanover embraced an idea to craft a drug that could selectively pass through the relatively leaky vasculature of cancer cells, pointing out that a similar



Ciechanover’s passion for science is evident.

strategy was used to develop Caelyx/Doxil — a liposome-encapsulated form of the chemotherapeutic agent doxorubicin. Because doxorubicin is highly toxic to the heart, it needs to be wrapped in an artificial liposome to reduce cardiotoxicity.

At each stage of the discussion, Ciechanover pushed the student researchers to think about implementation — to consider potential side effects or other problems that might undermine a promising idea. He pointed out that many strategies have already been worked out for overcoming hurdles in drug development, and suggested that young researchers would be well-advised to keep abreast of the approaches that have succeeded in the past.

The student presenters were energized by the chance to connect intellectually with a Nobel laureate. “It was more than one graduate student can ask for,” Anant said afterward. It was an honour, agreed Negi, who said that Ciechanover’s enthusiasm and optimism were contagious, spurring her to think more about adapting existing techniques to treat other human diseases. Ciechanover, she said, made the young researchers in the audience “think practically about solutions”.

The students were inspired by the fluid, informal conversation. “It was very intense scientifically, and a very sharp discussion,” said Gigin Lin of the Institute for Cancer Research in London. Christian Schulz, a PhD student in biochemistry at the University of Göttingen, Germany, agreed, adding that “it’s something different for the students to get to speak up more”.

In the end, despite the allusion to Renaissance pedagogy, the session did not closely resemble a masterclass where student and teacher engage in a one-on-one instruction. But the principle of an apprentice learning by example rather than through explicit instruction emerged nonetheless. And the arc of the discussion — from the intricacies of the ubiquitin system to the highest hurdles in drug development — provided a worthy intellectual blueprint for a young scientist to follow. Ideally, “the mentor doesn’t tell you what the next experiment should be,” Ciechanover said. But, by observing how the mentor approaches problems and thinks about solutions, his mindset becomes embedded in your own. ■

Kat McGowan is a journalist based in New York City.

PRACTICAL ADVICE FROM THE LAUREATES

Reality of life in the lab

Just because you’ve won a Nobel prize doesn’t exempt you from the everyday realities of a life in science. In Lindau’s Turning the Tables symposium, featuring a panel of Nobel laureates and young researchers, the laureates had an opportunity to query the students about their interests and plans. In the conversations that ensued, some elite names in science provided some down-to-earth advice:

Work hard: In response to a query from Thomas Steitz, who won the Nobel Prize in Chemistry in 2009, student panellists said they routinely worked long days, but didn’t always go to the lab on weekends. Steitz remarked that weekend working was common in his day.

Never eat lunch alone. Steitz lauded the late 1960s culture of the Laboratory of Molecular Biology in Cambridge, United Kingdom, where meals and tea breaks were always social events, leading to new ideas and collaborations. “Now, people eat their lunch in their office doing email, which I think is a lousy way of doing science.”

Speak up. Engage with your advisor — even one with a hands-off management style. “The least successful students and postdocs are the ones who are silent,” said Steitz.



Students and laureates turn the tables on each other.

Creativity is fed by diverse sources. “The best way for the brain to work is to be exposed to different things,” said Torsten Wiesel, who shared the 1981 Nobel Prize in Physiology or Medicine. Wiesel cited his life-long interest in art as an inspiration. “I see colleagues who work very hard, doing all the trivial work. Their lives miss some quality of joy.”

Targeted gene correction of α_1 -antitrypsin deficiency in induced pluripotent stem cells

Kosuke Yusa^{1*}, S. Tamir Rashid^{2,3*}, Helene Strick-Marchand^{4,5}, Ignacio Varela⁶, Pei-Qi Liu⁷, David E. Paschon⁷, Elena Miranda^{3,8}, Adriana Ordóñez³, Nicholas R. F. Hannan², Foad J. Rouhani^{1,2}, Sylvie Darche^{4,5}, Graeme Alexander⁹, Stefan J. Marciniak³, Noemi Fusaki^{10,11}, Mamoru Hasegawa¹⁰, Michael C. Holmes⁷, James P. Di Santo^{4,5}, David A. Lomas^{3*}, Allan Bradley^{1*} & Ludovic Vallier^{2*}

Human induced pluripotent stem cells (iPSCs) represent a unique opportunity for regenerative medicine because they offer the prospect of generating unlimited quantities of cells for autologous transplantation, with potential application in treatments for a broad range of disorders^{1–4}. However, the use of human iPSCs in the context of genetically inherited human disease will require the correction of disease-causing mutations in a manner that is fully compatible with clinical applications^{3,5}. The methods currently available, such as homologous recombination, lack the necessary efficiency and also leave residual sequences in the targeted genome⁶. Therefore, the development of new approaches to edit the mammalian genome is a prerequisite to delivering the clinical promise of human iPSCs. Here we show that a combination of zinc finger nucleases (ZFNs)⁷ and *piggyBac*^{8,9} technology in human iPSCs can achieve biallelic correction of a point mutation (Glu342Lys) in the α_1 -antitrypsin (*A1AT*, also known as *SERPINA1*) gene that is responsible for α_1 -antitrypsin deficiency. Genetic correction of human iPSCs restored the structure and function of *A1AT* in subsequently derived liver cells *in vitro* and *in vivo*. This approach is significantly more efficient than any other gene-targeting technology that is currently available and crucially prevents contamination of the host genome with residual non-human sequences. Our results provide the first proof of principle, to our knowledge, for the potential of combining human iPSCs with genetic correction to generate clinically relevant cells for autologous cell-based therapies.

At present, available methods for gene targeting rely on positive selection to isolate rare clones that have undergone homologous recombination. To remove the unwanted selection cassettes, Cre/*loxP* or FLP/*FRT* recombination systems are used, which leave behind single *loxP* or *FRT* sites^{10,11}. These small ectopic sequences have the potential to interfere with transcriptional regulatory elements of surrounding genes¹², most of which are not fully characterized in the human genome. An alternative method to remove selection cassettes is to convert them into transposons. The most suitable transposon for this purpose is *piggyBac*, a moth-derived DNA transposon, which can transpose efficiently in mammalian cells including human embryonic stem (ES) cells^{9,13}. A remarkable feature of this mobile element is seamless excision, which enables the removal of transgenes flanked by *piggyBac* inverted repeats without leaving any residual sequences^{9,14}.

To explore the use of *piggyBac* for the correction of point mutations, we designed a vector to correct an albino mutation (G290T substitution in the *Tyr* gene) in mouse iPSCs isolated from fibroblasts of the C57Bl6-*Tyr*^{c-Brd} strain¹⁵. The targeting vector was constructed, carrying a wild-type 290G sequence and a *PGK-puroAtk* cassette

flanked by *piggyBac* repeats into the TTAA site (Fig. 1a). After isolation of targeted clones, the selection cassette was excised from the mouse iPSC genome by transient expression of the *piggyBac* transposase and subsequent 1-(2-deoxy-2-fluoro- β -D-arabinofuranosyl)-5-iodouracil

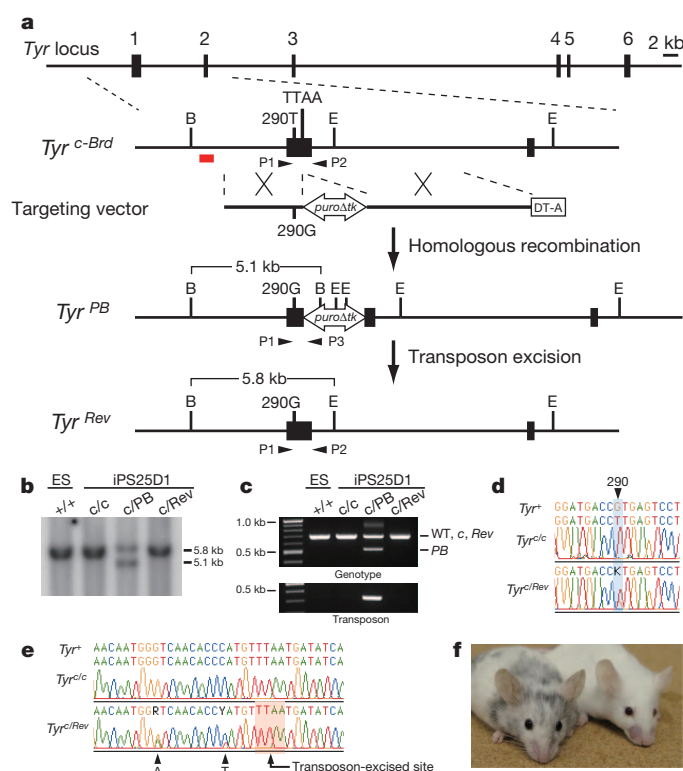


Figure 1 | Correction of the G290T mutation in the *Tyr* gene in mouse iPSCs. **a**, The strategy for precise genome modification using the *piggyBac* transposon. Top line, structure of the *Tyr* gene; red line, 5' external probe for Southern blot analysis; open arrow, *piggyBac* transposon carrying a *PGK-puroAtk* cassette; B, BamHI; E, EcoNI; P1, P2 and P3, PCR primers. **b**, **c**, Southern blot (**b**) and PCR analyses (**c**) showing insertion (c/PB) and excision (c/Rev) of the *piggyBac* transposon. ES, mouse ES cells as a control. **d**, **e**, Sequence analyses revealed correction of the G290T mutation (**d**) and seamless excision of the *piggyBac* transposon (**e**). Note that two silent mutations (A and T, indicated by arrowheads) introduced near the TTAA site were also detected. **f**, A chimaeric mouse generated by injecting corrected *Tyr*^{c/Rev} mouse iPSCs (left) shows black coat colour. Right, a non-injected albino mouse.

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²Anne McLaren Laboratory for Regenerative Medicine, Department of Surgery, West Forvie Building, Robinson Way, University of Cambridge, Cambridge CB2 0SZ, UK. ³Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB0 2XY, UK. ⁴Innate Immunity Unit, Institut Pasteur, 75724 Paris, France. ⁵INSERM, U668, 75724 Paris, France. ⁶Instituto de Biomedicina y Biotecnología de Cantabria (IBBT), CSIC-UC-SODERCAN Avda. Cardenal Herrera Oria s/n 39011 Santander, Spain. ⁷Sangamo BioSciences Inc., Richmond, California 94804, USA. ⁸Dept. Biología e Biotecnología 'Charles Darwin', Universidad de Roma 'La Sapienza', p.le Aldo Moro 5, 00185 Rome, Italy. ⁹Division of Gastroenterology and Hepatology, Department of Medicine, Cambridge University Hospitals NHS Trust, Cambridge CB2 2QQ, UK. ¹⁰DNAVEC Corporation, Tsukuba, Ibaraki 300-2611, Japan. ¹¹PRESTO, JST, Saitama 332-0012, Japan.

*These authors contributed equally to this work.

(FIAU) selection. Genomic modification was verified by Southern blot and polymerase chain reaction (PCR) analyses (Fig. 1b, c). The correction of the G290T mutation and seamless *piggyBac* excision were confirmed by sequence analyses (Fig. 1d, e). Two introduced silent mutations were observed, confirming that the T290G substitution was mediated by gene correction, not by spontaneous reversion (Fig. 1e). The function of the reverted allele was tested by injecting the corrected mouse iPSCs into albino mouse blastocysts. The resulting chimaeric mice had a black coat colour, indicating phenotypic correction of the albino mutation (Fig. 1f). These results collectively demonstrate that the *piggyBac* transposon can be used as a versatile tool for highly precise modification (for example, correction or mutation) of the mammalian genome at a single base-pair level.

We next explored whether this approach could be used to correct a mutation in human iPSCs derived from individuals with α_1 -antitrypsin deficiency (A1ATD)¹⁶. A1ATD is an autosomal recessive disorder found in 1 out of 2,000 individuals of North European descent and represents the most common inherited metabolic disease of the liver^{17,18}. It results from a single point mutation in the *A1AT* gene (the Z allele; Glu342Lys) that causes the protein to form ordered polymers within the endoplasmic reticulum of hepatocytes^{17,18}. The resulting inclusions cause cirrhosis for which the only current therapy is liver transplantation. The increasing shortage of donors and harmful effects of immunosuppressive treatments impose major limitations on organ transplantation, making the potential of human iPSC-based therapy highly attractive. Because homologous recombination is relatively inefficient in human ES cells⁶ we used ZFN technology, which stimulates gene targeting in human ES cells as well as human iPSCs^{7,10,19}. ZFN pairs were designed to specifically cleave the site of the Z mutation (Fig. 2a–c, Supplementary Table 1 and Supplementary Note). A targeting vector was constructed from isogenic DNA with *piggyBac* repeats flanking the *PGK-puroAtk* cassette (Fig. 2a). To minimize the distance between the mutation and the *piggyBac* transposon, a CTG leucine codon, 10-bp upstream of the mutation, was altered to a TTA leucine codon, generating the TTAA sequence, which would be left in the genome following *piggyBac* excision (Fig. 2b).

Puromycin-resistant human iPSC colonies obtained after co-electroporation of ZFN expression vectors and the targeting vector were screened for targeted clones by PCR. A1ATD-iPSC lines derived from three different patients yielded targeted clones (Table 1). Remarkably, 54% of the puromycin-resistant colonies were targeted on one allele, whereas 4% were the result of simultaneous targeting of both alleles (Supplementary Fig. 1).

To remove the *piggyBac*-flanked selection cassette from these modified clones, we transiently transfected two homozygously targeted clones (B-16 and C-64) with a hyperactive form of the *piggyBac* transposase⁸ and subjected them to FIAU selection. The genotype of the resulting FIAU-resistant colonies was analysed by PCR and confirmed by Southern blot (Fig. 2d and Supplementary Fig. 2a). Biallelic excision was observed in 11% of FIAU-resistant colonies (Table 2). Sequence analyses demonstrated that the Z mutation was corrected on both alleles and that transposon excision yielded a TTAA sequence as initially planned (Fig. 2b, e and Supplementary Fig. 2b). The resulting corrected iPSC lines maintained the expression of pluripotency markers for more than 20 passages and their abilities to differentiate into cells expressing markers of the three germ layers (Supplementary Fig. 3), indicating that genome modification did not alter the pluripotency of corrected human iPSCs.

Genomic instability is known to be associated with prolonged culture of human ES cells^{20,21} and mutations arising during genome modification would be another concern for clinical application of human iPSCs. Therefore, we analysed the genomic integrity of the human iPSC lines using comparative genomic hybridization (CGH) (Supplementary Table 2a–c). Two out of three A1ATD-iPSC primary lines differed from their parental fibroblasts, showing amplifications or deletions ranging from 20 kb to 1.3 Mb, including a gain of 20q11.21, a frequently

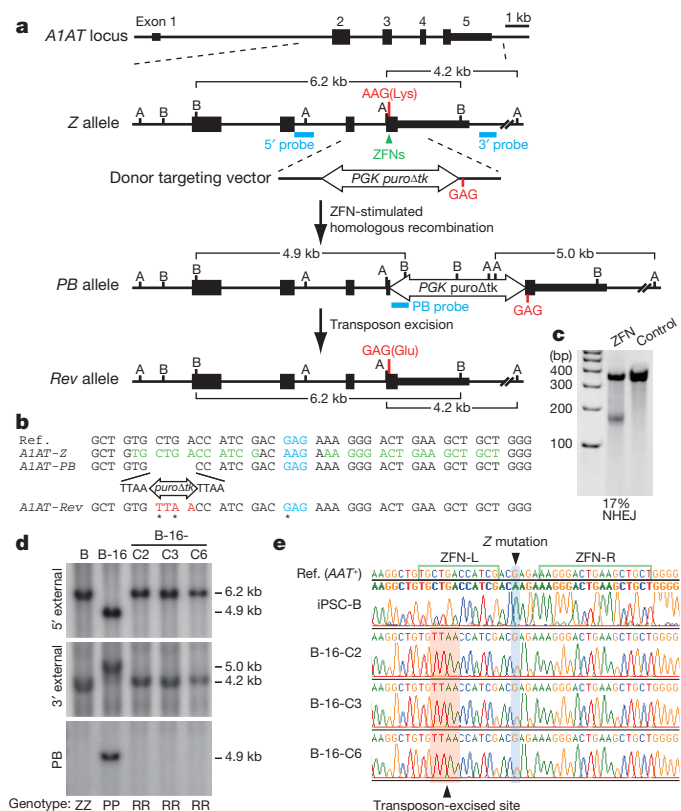


Figure 2 | Correction of the Z mutation in human A1ATD-iPSCs. **a**, The strategy for precise genome modification using ZFNs and the *piggyBac* transposon. Top line, structure of the *A1AT* gene; blue lines, Southern blot probes; thin and thick boxes, non-coding and coding exons, respectively; open arrow, *piggyBac* transposon; A, AflIII; B, BamHI. **b**, Sequences of wild-type (Ref.), Z, PB and Rev alleles. Amino acid position 342 (blue), recognition sites for ZFNs (green) and *piggyBac* excision site (red) are shown. Sequence changes in Rev allele from Z allele are indicated by asterisks. **c**, Surveyor nuclease assay showing the cleavage of the Z mutation in ZFN-transfected K562 cells. Non-transfected cells were used as a control. NHEJ, non-homologous end joining. **d**, Southern blot analysis showing biallelic *piggyBac* insertion (B-16) and biallelic excision (B-16-C2, -C3 and -C6) during correction of the A1ATD-iPSC line B. Genomic DNA was digested by BamHI (5' and PB probes) or AflIII (3' probe). Genotype: PP, homozygous for insertion of *piggyBac*; RR, homozygous for reverted allele; ZZ, homozygous for Z allele. **e**, Sequence analysis showing correction of the Z mutation in three corrected iPSC lines. Wild-type sequence (top line) and A1ATD-iPSC sequence (second line).

amplified region in human ES cells^{22,23} (see Supplementary Analysis and Supplementary Fig. 4). Line A retained a normal genome content compared to its parental fibroblast. Reassuringly, we found that after ZFN-stimulated targeting, four out of six homozygous clones had unaltered genomes compared to their parental iPSC lines. Sixteen cell

Table 1 | Summary of PCR genotyping of ZFN-stimulated gene targeting

A1ATD-iPSC line	Clones analysed	Het.*	Homo./Hemi.†	Het. + additional integrations‡	Homo./Hemi. + additional integrations‡	Non-targeted§
A	84	45	3	23	8	5
B	18	10	2	3	3	0
CII	216	112	9	52	21	22
Mean frequency (%)		54	6	23	12	5

* Het., clones heterozygous for PB allele.

† Homo./Hemi., clones homozygous or hemizygous for PB allele. Cells with one targeted allele and deletion of the other allele are indistinguishable from correctly targeted homozygous clones by PCR. Such cells are designated as hemizygotes.

‡ Vector backbone integration was analysed by PCR.

§ Clones showing incorrect PCR bands are included.

|| Sum of two independent experiments.

Table 2 | Frequencies of biallelic *piggyBac* excision

Cell line	Clones analysed	Biallelic excision without re-integration		Biallelic excision with re-integration	
		Number of clones	Frequency (%)	Number of clones	Frequency (%)
B-16	88	15	17	33	38
C-G4	94	5	5	19	20
Mean frequency (%)			11		29

lines with biallelic *piggyBac* excision were compared with their corresponding primary iPSCs and 12 had unaltered genomes. We also analysed the iPSC lines by SNP arrays to check for loss of heterozygosity and found that all lines analysed retained heterozygosity throughout their genome (Supplementary Fig. 5). This observation demonstrates that biallelic gene correction was the result of simultaneous homologous recombination followed by simultaneous excision at both alleles and that mitotic recombination was not involved in this process.

ZFN off-target cleavage and imprecise excision after multiple *piggyBac* transposition might introduce mutations into the genome. To investigate these possibilities at a single base-pair resolution, we sequenced exomes of the corrected B-16-C2 line and its parental fibroblast. Comparison of these exomes identified 29 mutations (Supplementary Table 3). The genesis of these mutations was determined by analysis of the primary iPSC line and the homozygously targeted intermediate. Twenty-four point mutations and one 1-bp deletion were detected in the primary iPSC line and four mutations arose during genetic correction: one during targeting and three during *piggyBac* excision. These mutations seemed to arise during culture as their genomic signatures were inconsistent with ZFN off-target sites or *piggyBac* integration sites (Supplementary Analysis). Taken together, we conclude that the combination of ZFNs with *piggyBac* provides a new method for rapid and clean correction of a point mutation in human iPSCs without affecting their basic characteristics.

To confirm that the genetic correction of A1ATD-iPSCs resulted in the expected phenotypic correction, iPSCs were differentiated *in vitro* into hepatocyte-like cells, the main cell type affected by the disease A1ATD. Differentiation of the corrected lines occurred as expected, resulting in a near homogenous population of hepatocyte-like cells (Supplementary Fig. 6a–c). Remarkably, CGH analysis of differentiated cells showed that hepatic differentiation neither increases the number of genetic abnormalities nor selects for cells with abnormal karyotype (Supplementary Table 2d). The resulting cells shared key functional attributes of their *in vivo* counterparts including glycogen storage, low density lipoprotein (LDL)-cholesterol uptake, albumin secretion and cytochrome P450 activity (Supplementary Fig. 6d–g). Importantly, immunofluorescence and enzyme-linked immunosorbent assay (ELISA) both confirmed the absence of mutant polymeric A1AT in corrected iPSC-derived hepatocyte-like cells that instead efficiently secreted normal endoglycosidase-H-insensitive monomeric A1AT (Fig. 3a–d). In addition, secreted A1AT showed an enzymatic inhibitory activity that was comparable to that obtained from normal adult hepatocytes (Fig. 3e), thereby suggesting that physiological restoration of enzyme inhibitory activity could be achieved.

Lastly, the *in vivo* function of corrected iPSC-derived hepatocyte-like cells (B-C16-2 line) was assessed following transplantation into the liver of *Alb-uPA^{+/+};Rag2^{-/-};Il2rg^{-/-}* mice via intrasplenic injection. Livers harvested 14 days after injection were colonized by human cells identified using antibodies specific to human albumin and A1AT (Fig. 3f, g). These human hepatocyte-like cells were distributed throughout the liver lobes and were seen to be integrated into the existing mouse parenchyma (Fig. 3f, g). In addition, human albumin was detected in the serum of transplanted animals for at least 5 weeks (Fig. 3h), whereas no tumour formation was detected in any mice. Therefore, corrected iPSC-derived hepatocyte-like cells were able to colonize the liver *in vivo* and show functional activities characteristic of their human ES-cell-derived counterparts²⁴. Collectively these analyses

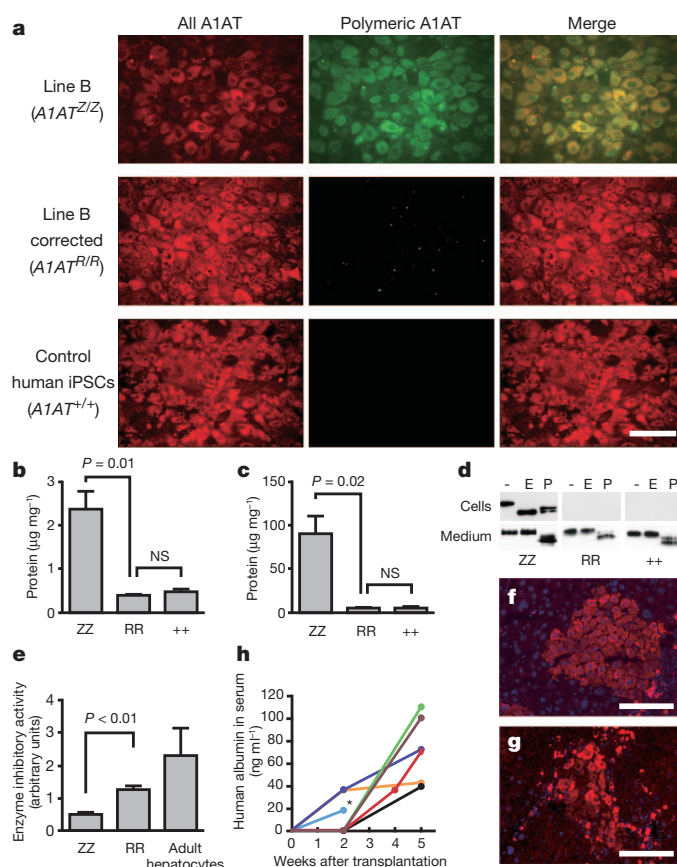


Figure 3 | Functional analysis of restored A1AT in corrected iPSC-derived hepatocyte-like cells. **a**, Immunofluorescence showing the absence of polymeric A1AT protein in hepatocyte-like cells generated from corrected iPSCs. All forms of A1AT (left panels) and misfolded polymeric A1AT (middle panels) are shown. **b**, **c**, ELISA to assess the intracellular (**b**) and secreted (**c**) levels of polymeric A1AT protein in hepatocyte-like cells derived from A1ATD-iPSCs (ZZ), corrected iPSCs (RR) and control human iPSCs (++). NS, not significant. **d**, Endoglycosidase H (E) and peptide:N-glycosidase (P) digestion of A1AT immunoprecipitated from uncorrected (ZZ), corrected (RR) and control (++) human iPSC-derived hepatocyte-like cells (upper panels) and corresponding culture medium (lower panels). **e**, Chymotrypsin ELISA showing that corrected cells (RR) have A1AT enzymatic inhibitory activity that is superior to uncorrected cells (ZZ) and close to adult hepatocytes. **f**, **g**, Immunofluorescence of transplanted liver sections detecting human albumin (**f**) and A1AT (**g**). DNA was counterstained with 4',6-diamidino-2-phenylindole (DAPI). **h**, ELISA read-out of human albumin in the mouse serum longitudinally followed for each mouse. Asterisk, the mouse was subjected to histology analysis. Scale bars, 100 μ m. Data in **b**, **c** and **e** are shown as mean \pm s.d. ($n = 3$). Student's *t*-test was performed.

demonstrate that genetic correction of the Z mutation resulted in functional restoration of A1AT in patient-derived cells.

All the experimental evidence described earlier strongly supports the applicability of genetic correction in patient-specific iPSCs for cell-based therapy of A1ATD. We therefore repeated the genetic correction in more clinically relevant cells using patient-specific iPSCs reprogrammed from fibroblasts with Sendai virus vectors, an integration-free method²⁵ (Supplementary Fig. 7a–f). One primary human iPSC line with an intact genome by CGH analysis (Supplementary Fig. 7e and Supplementary Table 4) was corrected by the method described earlier. The final product, iPSC-3-G5-A7, had the corrected A1AT, an intact genome compared to the parental fibroblast and expressed normal A1AT protein when differentiated to hepatocyte-like cells (Supplementary Fig. 8 and Supplementary Table 4). This is the first demonstration, to our knowledge, of the generation of mutation-corrected patient-specific iPSCs, which could realize the therapeutic promise of human iPSCs.

Here we demonstrate that ZFNs and *piggyBac* transposons enable simultaneous biallelic correction of diseased human iPSCs. No residual ectopic sequences remain at the site of correction and the genome seems to be undisturbed elsewhere. Although we could readily obtain cell lines without large genomic alterations during genetic modification, the resulting corrected human iPSCs carry 29 mutations in protein-coding exons, of which 22 were non-synonymous or splice site mutations. The probable impact of this mutation load needs to be considered in the context of the likely functional impact of the mutations, taking into account the normal germline load, accumulated somatic variation, the presence of compensating normal gene copies and the requirement for the gene product in the derived differentiated cells. From this point of view, only eight mutations might affect gene functions in hepatocyte-like cells (Supplementary Table 3). Nevertheless, the corrected iPSCs could efficiently differentiate to hepatocyte-like cells and engraft into the animal model for liver injury without tumour formation. Therefore, limited genomic abnormalities might have restricted biological consequences. Careful screening of primary and corrected human iPSCs using deep sequencing analyses would contribute to the safe use of human iPSCs in clinical applications.

iPSCs derived from different patients were effectively corrected, demonstrating that this method could be applied to a large number of A1ATD-iPSC lines. Because the biallelic correction could be carried out in less than 4 months, our approach may be compatible with large-scale production of corrected patient-specific iPSCs not only for A1ATD but also for other monogenic disorders.

METHODS SUMMARY

A1ATD-iPSCs were described previously¹⁶. 2×10^6 human iPSCs were co-transfected with ZFN expression vectors and the donor template, and subjected to puromycin selection ($1 \mu\text{g ml}^{-1}$) initiated 4 days after transfection. For transposon excision, targeted cells were transfected with pCMV-hyPB⁸, cultured for 4 days, re-plated and selected in 250 nM FIAU. To increase clonogenicity, cells were treated with ROCK inhibitor²⁶, Y-27632 ($10 \mu\text{M}$) 4 h before dissociation and 24 h after plating. Resulting colonies were picked 2 weeks later, analysed by PCR and further verified by Southern blot analysis. Primer sequences are listed in Supplementary Table 5.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 January; accepted 8 August 2011.

Published online 12 October 2011.

1. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
2. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
3. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **24**, 2239–2263 (2010).
4. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920–1923 (2007).
5. Fairchild, P. J. The challenge of immunogenicity in the quest for induced pluripotency. *Nature Rev. Immunol.* **10**, 868–875 (2010).
6. Tenzen, T., Zembowicz, F. & Cowan, C. A. Genome modification in human embryonic stem cells. *J. Cell. Physiol.* **222**, 278–281 (2010).
7. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature Rev. Genet.* **11**, 636–646 (2010).
8. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive *piggyBac* transposase for mammalian applications. *Proc. Natl Acad. Sci. USA* **108**, 1531–1536 (2011).
9. Wang, W. *et al.* Chromosomal transposition of *PiggyBac* in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 9290–9295 (2008).
10. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnol.* **27**, 851–857 (2009).
11. van der Weyden, L., Adams, D. J. & Bradley, A. Tools for targeted manipulation of the mouse genome. *Physiol. Genomics* **11**, 133–164 (2002).

12. Meier, I. D. *et al.* Short DNA sequences inserted for gene targeting can accidentally interfere with off-target gene expression. *FASEB J.* **24**, 1714–1724 (2010).
13. Lacoste, A., Berenshteyn, F. & Brivanlou, A. H. An efficient and reversible transposable system for gene delivery and lineage-specific differentiation in human embryonic stem cells. *Cell Stem Cell* **5**, 332–342 (2009).
14. Fraser, M. J., Ciszczon, T., Elick, T. & Bauser, C. Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol. Biol.* **5**, 141–151 (1996).
15. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the *piggyBac* transposon. *Nature Methods* **6**, 363–369 (2009).
16. Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120**, 3127–3136 (2010).
17. Perlmutter, D. H. Autophagic disposal of the aggregation-prone protein that causes liver inflammation and carcinogenesis in α -1-antitrypsin deficiency. *Cell Death Differ.* **16**, 39–45 (2009).
18. Goopu, B. & Lomas, D. A. Conformational pathology of the serpins: themes, variations, and therapeutic strategies. *Annu. Rev. Biochem.* **78**, 147–176 (2009).
19. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97–110 (2009).
20. Mitalipova, M. M. *et al.* Preserving the genetic integrity of human embryonic stem cells. *Nature Biotechnol.* **23**, 19–20 (2005).
21. Baker, D. E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nature Biotechnol.* **25**, 207–215 (2007).
22. Lefort, N. *et al.* Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nature Biotechnol.* **26**, 1364–1366 (2008).
23. Spits, C. *et al.* Recurrent chromosomal abnormalities in human embryonic stem cells. *Nature Biotechnol.* **26**, 1361–1363 (2008).
24. Touboul, T. *et al.* Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51**, 1754–1765 (2010).
25. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **85**, 348–362 (2009).
26. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnol.* **25**, 681–686 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Klug and M. Minczuk for their advice, M. A. Li for comments on the manuscript, P. Ellis, N. Hammond and C. McGee for CGH analysis, the Sanger Institute sequencing facility for exome sequencing, N. Conte and S. Rice for assistance with bioinformatic analysis, M. Alexander for her help with cell culture reagents. We also thank L. Zhang, S. Hinkley and the production group for ZFN assembly and validation, K. Tong and X. Meng for technical assistance, J. C. Miller and E. Leung for ZFN off-target site analysis and S. Abrahamson and P. D. Gregory for careful reading of the manuscript. This work was supported by the Wellcome Trust (WT077187; A.B.), the MRC Senior non-clinical fellowship and the Cambridge Hospitals National Institute for Health Research Biomedical Research Center (L.V.), the Medical Research Council and Papworth NHS Trust (D.A.L.), the Bill and Melinda Gates Foundation, Inserm and Institut Pasteur (H.S.-M.) and Japan Science and Technology Agency (N.F.). K.Y. is supported by a postdoctoral fellowship of Japan Society for the Promotion of Science. S.T.R. and F.J.R. are Wellcome Trust Clinical Training Fellows. I.V. is supported by a fellowship from the International Human Frontiers Science Program Organization.

Author Contributions K.Y. and S.T.R. are joint first authors. D.A.L., A.B. and L.V. contributed equally to this work. K.Y., S.T.R., D.A.L., A.B. and L.V. conceived the research and wrote the manuscript with comments from all authors. K.Y. performed gene correction in mouse and human iPSCs and conducted all experiments using *piggyBac* in Cambridge, UK. S.T.R., E.M., A.O., N.R.F.H., F.J.R., G.A. and S.J.M. performed *in vitro* phenotypic analysis of corrected human iPSCs. S.T.R., H.S.-M., S.D. and J.P.D.S. performed *in vivo* work. I.V. performed data analysis of exome sequencing. P.Q.-L., D.E.P. and M.C.H. generated and validated ZFNs. N.F. and M.H. generated Sendai virus vectors.

Author Information Exome sequence data have been deposited at the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>) hosted by the European Bioinformatics Institute under accession EGAS00001000055. CGH and SNP array data have been deposited with EBI ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MEXP-3316 and with Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE31035, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.B. (abradley@sanger.ac.uk) or L.V. (lv225@cam.ac.uk).

METHODS

Plasmid construction. Gateway-adapted *piggyBac* transposon vectors: a destination vector pPB-R1R2-NP was constructed as follows. The *attR1* and *attR2* sites were PCR-generated and digested by *NheI*/*HindIII* and *XhoI*/*SpeI*, respectively. *EM7-neo* was PCR-generated and digested by *HindIII*/*XhoI*. These three fragments were then cloned into the *NheI*-*SpeI* site of pPB-LR5 (ref. 27), resulting in pPB-R1R2-Neo. An *EcoRI*-*XbaI* fragment containing *PheS* was excised from pR6K-R1R2-ZP²⁸, blunt-ended and cloned into the blunt *XhoI* site of pPB-R1R2-Neo, resulting in pPB-R1R2-NP. An entry vector pENTR-PGK*puroAtk* was constructed by cloning a *KpnI*-*NotI* *PGK-puroAtk* fragment into the *KpnI*-*NotI* site of pENTR-2B.

A targeting vector for *Tyr*: the targeting vector was constructed using BAC recombining. A bacterial artificial chromosome (BAC) clone RP24-221M7 was introduced into *Escherichia coli* strain EL350 (ref. 29). A mini targeting vector was first constructed to modify the *Tyr* gene on the BAC. Left and right homology arms were PCR-generated and digested by *AscI*/*BsiWI* and *NsiI*/*PacI*, respectively. The transposon fragment was excised from pPB-R1R2-NP by *NsiI*/*BsiWI* digestion. These three fragments were then cloned into the *AscI*/*PacI* site of pMCS, resulting in pMCS-Tyr-NP. An *AscI*-*PacI* fragment was excised from pMCS-Tyr-NP and used for BAC targeting. A retrieving vector was constructed by cloning PCR-generated left and right homology arms into the *XhoI*/*AscI* site of pMCS-DTA, following *AscI*/*HindIII* and *XhoI*/*HindIII* digestion of the left and right arm, respectively. The retrieving vector was linearized by *HindIII* digestion and used to retrieve the 3.0-kb 5' arm, the transposon and the 6.5-kb 3' arm. Lastly, the *Neo-PheS* cassette was replaced with the *PGK-puroAtk* cassette by Gateway cloning, resulting in pDTA-Tyr^{NP}. The targeting vector was linearized by *AscI* before electroporation into the albino iPSCs.

A donor template vector for *A1AT*: a 2-kb fragment, which contained 1 kb on both sides of the Z mutation, was first PCR-amplified using genomic DNA from A1ATD-iPSC line B as a template and cloned into pCR4-blunt-TOPO (Invitrogen), resulting in pCR4-AAT_Z. To construct a donor template with corrected sequence and a *piggyBac* transposon, the 5' arm and 3' arm were PCR-amplified and digested with *AscI*/*NsiI* and *BsiWI*/*PacI*, respectively. The *NsiI*-*BsiWI* fragment containing a *piggyBac* transposon with the *Neo-PheS* cassette was excised from pPB-R1R2-NP. The digested fragments were cloned into the *AscI*-*PacI* site of pMCS, resulting in pMCS-AAT-PB:NP. The *Neo-PheS* cassette was subsequently replaced with a *PGK-puroAtk* cassette by Gateway cloning, resulting in the final donor vector, pMCS-AAT-PB:PGK*puroAtk*.

The plasmids (pPB-R1R2-NP, pENTR-PGK*puroAtk*, pMCS-AAT-PB:PGK*puroAtk*) have been deposited in the Wellcome Trust Sanger Institute Archives and are available upon request (<http://www.sanger.ac.uk/technology/clonerequests/>).

Cell culture. Appropriate ethical approval and patient consent were obtained (Ethics reference no. 08/H0311/201; R&D no. A091485). A1ATD-iPSCs (ref. 16; A, patient 2 line 1; B, patient 1 line 1; C, patient 3 line 1) were cultured on mouse embryonic fibroblast (MEF)-feeder layers in human ES cell medium: DMEM/F12 supplemented with 20% knockout serum replacement, 1 mM GlutaMax, 0.1 mM 2-mercaptoethanol, 1× non-essential amino acids and 4 ng ml⁻¹ FGF2 (Invitrogen). Subculture was performed every 5–7 days by detaching human iPSCs by incubation in 0.5 mg ml⁻¹ dispase and 0.5 mg ml⁻¹ collagenase type IV for 1 h at 37 °C, collecting detached human iPSC colonies, breaking down into small clumps and plating them onto new feeder plates. MEFs (CF1 or B6129F1) were cultured in DMEM containing 10% FCS, 2 mM glutamine, 0.1 mM 2-mercaptoethanol and 1× non-essential amino acids. Mouse iPSCs (iPS25A1; ref. 15) were cultured on MEF-feeder layers in mouse ES cell medium: KO-DMEM supplemented with 15% FBS, 1 mM GlutaMax, 0.1 mM 2-mercaptoethanol, 1× non-essential amino acids and 1,000 unit ml⁻¹ LIF (Millipore).

Gene targeting and transposon excision in mouse iPSCs. 1 × 10⁷ cells were electroporated with 25 µg of a linearized targeting vector in 800 µl of HEPES-buffered saline using a Gene Pulser II electroporator (230 V, 500 µF) and plated onto three 10-cm dishes. The next day, puromycin selection (1 µg ml⁻¹) was initiated. Resulting colonies were picked and screened by PCR. Targeted clones were expanded and further verified by Southern blot analysis. Correctly targeted clones were then subjected to transposon excision. 2 × 10⁶ cells were electroporated with 40 µg of pCMV-hyPBase in 800 µl of HEPES-buffered saline using a Gene Pulser II electroporator (230 V, 500 µF) and plated onto one well of a 6-well plate. After passage once, cells were replated on day 4 at 5 × 10⁵ cells per 10-cm dish. On the following day, FIAU (0.2 µM) selection was initiated. On day 5 of selection, FIAU was withdrawn. Resulting colonies were picked at day 7 and screened by PCR. Primer sequences to detect homologous recombination are shown in Supplementary Table 5.

ZFN-mediated gene targeting in A1ATD-iPSCs. On the day of electroporation (day 0), near-confluent cells were pre-treated with a ROCK inhibitor²⁶ (Y-27632,

Sigma) at 10 µM for 3–4 h before electroporation. Cells were then washed with PBS once, detached by Accutase (Millipore; 10 min at 37 °C) and mixed with DMEM/F12 containing 10% FCS. Cells were dissociated into single-cell suspension by vigorous pipetting and counted. 2 × 10⁶ cells were pelleted and mixed with 5 µg of a 5'-ZFN expression vector, 5 µg of a 3'-ZFN expression vector and 2 µg of the donor template in 100 µl of human ES cell solution 1 (Lonza). The cell suspension was transferred to a cuvette and electroporated using the Amaxa Nucleofector device (Lonza) with program A23. The electroporated cells were plated onto one or two 10-cm feeder dishes in MEF-conditioned human ES cell medium containing 10 µM Y-27632. Human ES cell medium without any drug was used for daily medium change between days 1–3. On day 4, puromycin selection (1 µg ml⁻¹) was started. On day 6, medium was changed to MEF-conditioned human ES cell medium containing 0.5 µg ml⁻¹ puromycin, which was used for medium change at every other day until picking colonies. Resulting colonies were picked on day 13–17. Colonies were cut into two pieces. One half was transferred onto one well of a 24-well feeder plate and the other half was lysed and used for PCR genotyping. PCR-positive clones were further expanded and homologous recombination was verified by Southern blot analysis.

Transposon excision in homozygously targeted human iPSCs. Homozygously targeted clones (B-16, C-G4, SeV-1-C3 and SeV-3-G5) were used for transposon removal. Line-A-derived clones were omitted because this line showed a lower capability of differentiating into endodermal lineages. Cells were prepared as described earlier. 2 × 10⁶ cells were mixed with 10 µg of the hyperactive *piggyBac* transposase expression vector (pCMV-hyPBase⁸) in 100 µl of human ES cell solution 1 and electroporated using the Nucleofector device with the program A23. Electroporated cells were plated onto a 6-well plate in 1:2, 1:3 and 1:6 dilutions in MEF-conditioned human ES cell medium containing 10 µM Y-27632. Note that ROCK inhibitor was added to the culture medium until day 6 in this experiment. On day 2, cells with ~80% confluency were passaged using Accutase at a split ratio of 1:2, 1:3 and 1:6 into 6-well plates. On day 4, cells with ~80% confluency were washed with PBS, detached with Accutase, suspended in human ES cell medium and pelleted. Cells were resuspended in human ES cell medium into single-cell level and counted. 1 × 10⁴ cells were then plated onto one 10-cm dish in human ES cell medium containing 10 µM Y-27632. 16–18 h after plating (day 5), medium was changed to human ES cell medium containing 0.25 µM FIAU and 10 µM Y-27632. On day 6, medium was changed to human ES cell medium containing 0.25 µM FIAU and then medium was changed every other day. Genotype and deletion of the *piggyBac* transposon were analysed by PCR and further verified by Southern blot analysis.

CGH analysis. Genomic DNA was extracted using a DNeasy kit (Qiagen). Agilent 244K human genome arrays were used following the manufacturer's protocol. The arrays were scanned with an Agilent microarray scanner and data were generated by Agilent Feature Extraction software. CGH calls were made with Agilent's DNA analytics software using the ADM2 algorithm (6.0 threshold) with a minimum of 5 probes in the region as a filter.

SNP analysis. An Illumina HumanCytoSNP-12 SNP array was used following the manufacturer's protocol. Genotype calls were performed by Illumina's GenomeStudio. B allele frequency and log R ratio were analysed by KaryoStudio. CNVpartition v2.4.4 bundled in KaryoStudio was used for copy number analysis.

ZFN design. ZFNs were designed against a region containing the Z mutation in the *A1AT* gene (see Fig. 2a, b) and assembled as previously described³⁰. The amino acid residues at positions '−1' to '+6' of the recognition α-helix^{31,32} of each of the zinc finger DNA-binding domains for each DNA triplet target are shown in Supplementary Table 2. The ZFNs were linked to the wild-type *FokI* catalytic domain. The activity of the ZFN at the endogenous target site was determined using the Surveyor Nuclease assay as previously described³³.

Human iPSC-derived hepatocyte-like cell transplantation. All mice were housed in pathogen-free conditions and animal studies were approved by the committee on animal experimentation of the Institut Pasteur and by the French Ministry of Agriculture. Differentiated cells (5 × 10⁵ cells per animal in 50 µl DMEM) were injected into the spleens of 3- to 4-week-old *Alb-uPA*^{+/+}; *Rag2*^{−/−}; *Il2rg*^{−/−} mice (*n* = 7). The recipient mouse was killed 2 weeks after transplantation for histological analysis. Blood samples were collected and human albumin in plasma was quantified by ELISA (Bethy Laboratories). Frozen liver sections were analysed by immunofluorescence with human albumin (Dako) or human A1AT (Dako) specific antibodies. Non-transplanted mice were used as controls.

Exome sequencing. The corrected iPSC line, B-16-C2, and its parental fibroblasts were analysed. Exome sequencing and analysis were performed as described previously³⁴ with minor modifications. Exome pull-down was performed using an Agilent SureSelect Human All Exon 50Mb Kit according to the manufacturer's instructions. Enriched DNA was sequenced on an Illumina HiSeq 2000 (75-bp paired-end sequencing). 90.32% (Fibroblast-B) and 90.72% (B-16-C2) of total targeted regions were covered with more than 10× sequencing depth,

covering 93.01% and 93.35% of CCDS exons, respectively. Substitutions in the coding sequence were called as positions with at least 20% of reads reporting a different base with respect the reference human sequence (GRCh37). Additionally, somatic mutations were identified by comparing the sequence with the control fibroblasts, and removing the common polymorphisms described in dbSNP and in the 1000 Genomes Project³⁵. Small insertions and deletions were identified using samtools, as the ones not present in the control cell line and that had at least 20× of coverage and 20% of the reads reporting the mutation. Validation of mutations was carried out by Sanger capillary sequencing on parental Fibroblast-B, A1ATD-iPSC line B, the homozygously targeted B-16 cells and the *piggyBac*-excised B-16-C2 cells.

Other experimental procedures. Sendai virus regrogramming, RT-PCR, quantitative RT-PCR, bisulphite sequencing, immunostaining, flow cytometric analysis, ELISA and EndoH analysis were performed as described previously^{16,24,25,36}.

27. Cadinanos, J. & Bradley, A. Generation of an inducible and optimized *piggyBac* transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
28. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
29. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
30. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
31. Beerli, R. R. & Barbas, C. F. III Engineering polydactyl zinc-finger transcription factors. *Nature Biotechnol.* **20**, 135–141 (2002).
32. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
33. Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
34. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
35. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
36. Seki, T. *et al.* Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* **7**, 11–14 (2010).

Feedback from rhodopsin controls *rhodopsin* exclusion in *Drosophila* photoreceptors

Daniel Vasiliasauskas¹, Esteban O. Mazzoni^{1†}, Simon G. Sprecher^{1†}, Konstantin Brodetskiy¹, Robert J. Johnston Jr¹, Preetmoninder Lidder¹, Nina Vogt¹, Arzu Celik^{1†} & Claude Desplan¹

Sensory systems with high discriminatory power use neurons that express only one of several alternative sensory receptor proteins. This exclusive receptor gene expression restricts the sensitivity spectrum of neurons and is coordinated with the choice of their synaptic targets^{1–3}. However, little is known about how it is maintained throughout the life of a neuron. Here we show that the green-light sensing receptor rhodopsin 6 (Rh6) acts to exclude an alternative blue-sensitive rhodopsin 5 (Rh5) from a subset of *Drosophila* R8 photoreceptor neurons⁴. Loss of Rh6 leads to a gradual expansion of Rh5 expression into all R8 photoreceptors of the ageing adult retina. The Rh6 feedback signal results in repression of the *rh5* promoter and can be mimicked by other *Drosophila* rhodopsins; it is partly dependent on activation of rhodopsin by light, and relies on G_{aq} activity, but not on the subsequent steps of the phototransduction cascade⁵. Our observations reveal a thus far unappreciated spectral plasticity of R8 photoreceptors, and identify rhodopsin feedback as an exclusion mechanism.

In the *Drosophila* visual system, rhodopsins (Rh), G-protein-coupled receptors, detect light and initiate the phototransduction cascade leading to depolarization of photoreceptor neurons⁵. Each ommatidium, the unit eye of the adult retina, contains eight photoreceptors. Six outer photoreceptors, R1–R6, express Rh1 and are involved in motion detection and dim light vision (reviewed in ref. 4). Inner photoreceptors R7 and R8 mediate colour vision and define two main ommatidial subtypes based on the rhodopsins they express: in pale (p) ommatidia, pR7 expresses ultraviolet-sensitive Rh3 whereas pR8 expresses Rh5; in yellow (y) ommatidia, yR7 expresses a distinct ultraviolet-sensitive Rh4 whereas yR8 expresses Rh6⁴. Subtypes p and y are distributed stochastically throughout the main part of the retina with an approximate 30:70 ratio (Fig. 1c)⁶. An exception to the exclusive rhodopsin expression exists in the medio-dorsal area of the eye, where although the p and y subsets are correctly specified, Rh3/Rh4 are co-expressed in yR7s⁷. This rhodopsin expression pattern is established by a well-understood developmental program executed during pupal stages^{4,8,9} (Supplementary Fig. 1a, b).

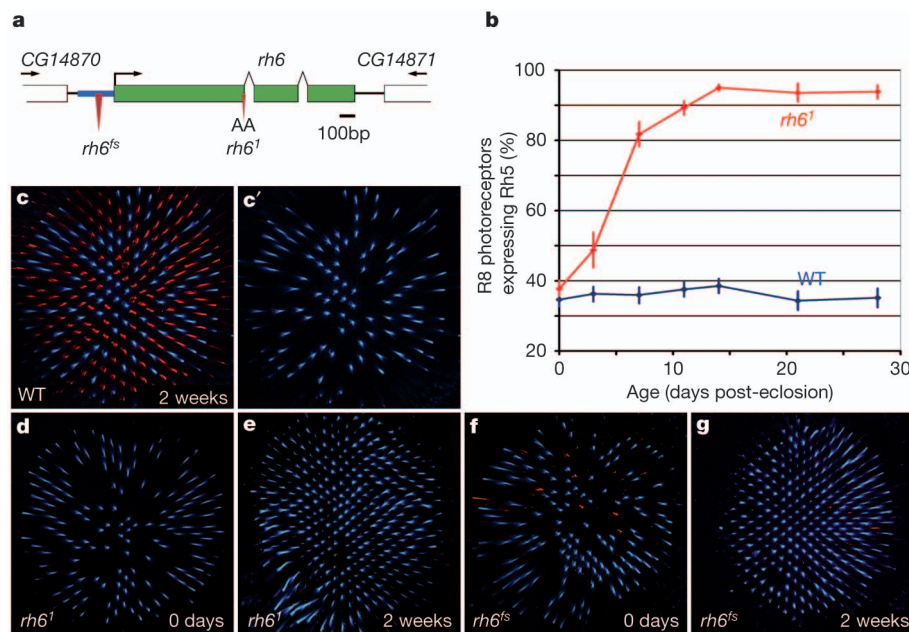


Figure 1 | Rh6 acts to repress Rh5 expression in yR8 photoreceptors.

a, Genomic *rh6* locus. The promoter region sufficient to drive *rh6* expression in yR8 is in blue, exons are in green and mutations in red. In *rh6*^{fs} mutants, 58 bp of the promoter are deleted. In *rh6*¹ mutants, 21 bp at the first exon–intron junction are replaced with AA, leading to an immediate truncation of the open reading frame. **b**, Percentage of R8 photoreceptors expressing Rh5 as a function of time (days post-eclosion) in wild-type (WT, blue) and *rh6*¹ mutants (red). Error bars, 84% confidence intervals. **c–g**, Wholemount retinas stained with

specific antibodies for Rh5 (blue) and Rh6 (red). **c, c'**, Normal expression of Rh5 and Rh6 in 2-week-old flies. **c'**, Rh5 alone. **d, e**, In *rh6*¹ mutants, Rh5 is gradually de-repressed. At eclosion, retinas have a normal number of Rh5-expressing R8s (**d**). By 2 weeks post-eclosion, most R8s express Rh5 (**e**). **f, g**, *rh6*^{fs} promoter mutation leads to loss of detectable Rh6 expression in almost all yR8s. As in *rh6*¹ mutants (**d, e**), at eclosion *rh6*^{fs} retinas have a normal number of Rh5-expressing R8s (**f**), but by 2 weeks post-eclosion, most R8 express Rh5 (**g**).

¹Center for Developmental Genetics, Department of Biology, New York University, New York, New York 10003, USA. [†]Present addresses: Departments of Pathology, Neurology, and Neuroscience, Columbia University Medical Center, 630 West 168 Street, New York, New York 10032, USA (E.O.M.); Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland (S.G.S.); Department of Molecular Biology and Genetics, Bogazici University, 34342 Bebek, Istanbul, Turkey (A.C.).

It is unknown, however, how p and y photoreceptor subtypes are maintained in the adult fly. The example of vertebrate olfaction, where sensory receptors act to repress expression of alternative receptor genes^{10–14}, led us to ask whether rhodopsins themselves participate in maintaining their mutual exclusion by analysing rhodopsin expression in various *rh* mutants. We found that in *rh6¹* mutants (Fig. 1a), the number of R8 cells expressing Rh5 increases dramatically and that this expansion of Rh5 is age dependent (Fig. 1b, d, e and Supplementary Table 1). In 1-day-old *rh6¹* mutant flies, Rh5 expression appears normal, with approximately 38% of R8s expressing uniformly high levels of Rh5 protein. In 3-day-old flies, additional R8s begin to express low levels of Rh5. By 14 days, nearly all (95%) R8s express Rh5. The levels of ectopic Rh5 in individual yR8s also increase over time, but remain variable and often are lower than in pR8 (Fig. 1e). In control flies, the number of Rh5-expressing R8s and the levels of expression remain stable as the flies age (36%, Fig. 1b, c and Supplementary Table 1).

We next asked if other rhodopsins are controlled by rhodopsin-mediated repression. We examined whether Rh6 expression is de-repressed in *rh5* mutants, but did not detect any Rh6 protein in pR8 in 3-week-old *rh5* mutants (Supplementary Fig. 2f). Expression of the non-R8 rhodopsins Rh1, Rh3 and Rh4 also remains normal in *rh5* or *rh6* mutants older than 3 weeks as well as in *rh5;rh6* double mutants (Supplementary Fig. 2a–e, g, h). Nonsense mutations in *rh3* or *rh4* genes do not affect expression of the remaining rhodopsins in R7s in either young or old (over 3 weeks) flies (Supplementary Figs 3 and 4, and data not shown). Thus, a rhodopsin-dependent mechanism for controlling rhodopsin expression occurs only in yR8s. Moreover, Rh5 is the only rhodopsin that is actively repressed by Rh6.

In the *rh6¹* allele, commonly found in laboratory stocks, a short deletion that spans the first exon–intron junction leads to a truncation of the protein after its fifth transmembrane domain¹⁵ (Fig. 1a). The levels of *rh6* messenger RNA (mRNA) measured by quantitative PCR with reverse transcription (qRT–PCR) are more than tenfold lower in *rh6¹* mutants than in wild-type flies (Supplementary Fig. 9a), probably because of nonsense-mediated decay. The Rh5 de-repression phenotype does not become more severe when *rh6¹* is placed over a deficiency (Supplementary Fig. 9b), suggesting that *rh6¹* is a null allele. Also, *rh6¹* can be rescued by a 2,575 base pair (bp) genomic fragment encompassing the *rh6* locus (Supplementary Figs 7a and 9b). Hereafter we refer to both *rh6¹* homozygotes and *rh6¹* trans-heterozygotes over a deficiency as *rh6* mutants and, unless otherwise noted, all phenotypes described are in ‘old’ flies 2 weeks post-eclosion or older.

We identified a second *rh6* allele, also in a laboratory stock, which we named *frank sinatra* (*rh6^{fs}*) after the singer known as “Ol’ blue eyes” (Fig. 1a). This mutation removes 58 bp of the *rh6* regulatory region without affecting the coding sequence. In *rh6^{fs}* mutants, Rh6 protein is detectable only in a few R8s in retinas of young flies ($6.5\% \pm 4.4$ SD, Supplementary Fig. 9b) where it is expressed at levels generally lower than normal (Fig. 1f, g). As in *rh6¹* mutants, Rh5 is initially expressed normally in 41% of R8 in *rh6^{fs}* flies (Fig. 1f), leaving most yR8s devoid of rhodopsin expression. However, Rh5 becomes broadly de-repressed in R8s of old flies (Fig. 1g and Supplementary Fig. 9b). Rh5 is rarely expressed in the few Rh6-positive R8s of *rh6^{fs}* mutants and co-expression only occurs in cells with low Rh6 levels (not shown). We also used a *rh6* promoter-based driver (*rh6-Gal4*) to express a RNA interference (RNAi) construct targeting *rh6* in differentiated yR8s. Although this does not completely abolish Rh6 in yR8 rhabdomeres, it leads to de-repression of Rh5 in old flies (Supplementary Fig. 5a). These results support the idea that reducing the levels of normal Rh6 activity leads over time to de-repression of Rh5 expression in yR8s.

Repression of Rh5 by Rh6 in wild-type yR8 could occur transcriptionally, or post-transcriptionally. We thus asked whether *rh5* mRNA expression is de-repressed in *rh6* mutants by performing *in situ* hybridization. *rh5* mRNA is present in many more R8s in old *rh6*

mutants than in age-matched wild-type flies (Fig. 2a, b). To visualize this phenotype more clearly, we repeated the experiment in a *sevenless* (*sev*) mutant background in which R7 photoreceptors are absent¹⁶. Because specification of *rh5*-expressing pR8s depends on the overlying pR7s (Supplementary Fig. 1a), most cells become yR8 and express Rh6 in *sev* flies while Rh5 is only expressed in a few R8 photoreceptors^{17–19} (~3%, Fig. 2c). However, in old *sev;rh6* double mutants, *rh5* mRNA is de-repressed in most R8s (Fig. 2d). We also quantified changes in *rh5* mRNA expression using qRT–PCR: in 2-week-old *rh6* mutants, *rh5* mRNA more than doubles over normal levels (Supplementary Fig. 9a). To ask whether this occurs through repression of the *rh5* promoter rather than by affecting mRNA stability, we analysed the expression of a *rh5* reporter (*rh5>GFP* (green fluorescent protein)) containing a –690 to +50 *rh5* promoter fragment²⁰. In control flies, *rh5>GFP* is co-expressed with Rh5 protein in pR8s (Fig. 2e). In *rh6* mutants *rh5>GFP* expression begins normally but with age expands to most yR8s (Fig. 2f and Supplementary Fig. 9f). This supports the model that Rh6 generates a feedback signal that acts to repress transcription from the *rh5* promoter and that the relevant regulatory sites are contained within the short promoter fragment of the *rh5>GFP* transgene.

Expression of *rh5* in yR8s of *rh6* mutants could be due to a change in yR8 cell identity, either during specification or in adults. To test this, we first asked whether a reporter for *rh6* expression (*rh6-lacZ*) is correctly activated in *rh6* mutant flies. In young *rh6* mutants, *rh6-lacZ* is robustly expressed in R8s in a pattern complementary to Rh5 expression (Fig. 3a and Supplementary Fig. 9c), suggesting correct specification of the yR8 subtype. As the fly ages, these cells de-repress Rh5 but remain positive for β -galactosidase (β Gal) (Fig. 3b and Supplementary Fig. 9c). We also tested for a possible yR8-to-pR8 fate transition using the marker genes that specify these cells. The p versus y subtype specification of R8 cells depends on an R8-intrinsic bistable

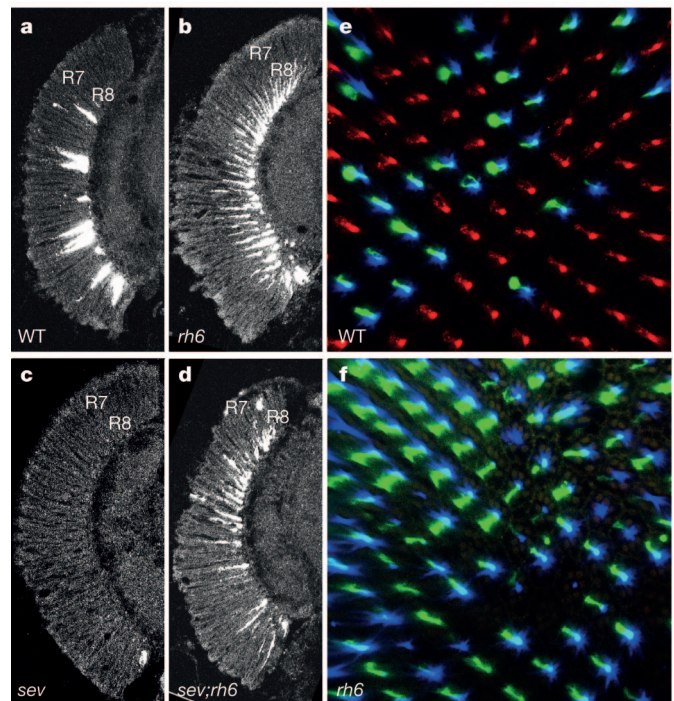


Figure 2 | Rh6 represses transcription of the *rh5* gene. a–d, *rh5* mRNA, detected by *in situ* hybridization in transverse cryo-sections of 3-week-old fly eyes. Many more cells are expressing *rh5* mRNA in the R8 layer of *rh6* mutants (b) compared with wild-type flies (a). In *sev* mutants, very few cells express *rh5* (c). However, in *sev;rh6* double mutants, *rh5* is extensively de-repressed in R8 photoreceptors (d). e, f, In 3-week-old control flies, a *rh5* reporter (*rh5>GFP*) (green) is expressed in pR8 cells that also express Rh5 protein (blue), but not in yR8 cells which express Rh6 (red) (e). In *rh6* mutants, *rh5>GFP* is de-repressed in most yR8 cells (f).

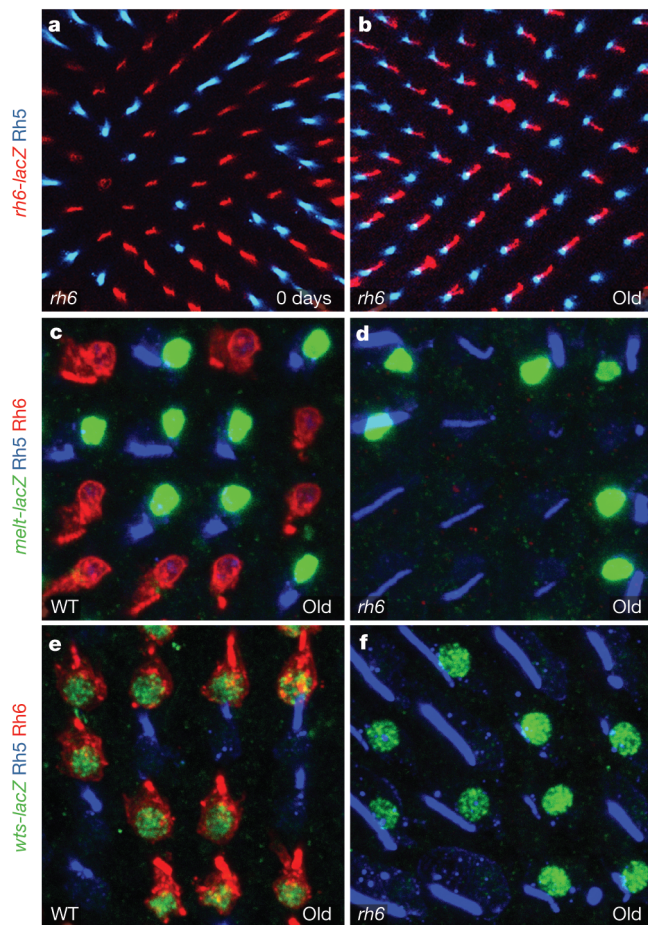


Figure 3 | Mutation of *rh6* does not lead to change in yR8 cell identity. **a, b,** A *rh6-lacZ* reporter (red) is expressed normally in *rh6* mutants. It is induced in a pattern complementary to the expression of Rh5 (blue) in young flies (**a**). In 2-week-old *rh6* mutants, Rh5 expression expands into the *lacZ*-positive, yR8 cells (**b**). **c–f,** Z-projections of confocal stacks encompassing nuclei and Rh-containing rhabdomeres of R8 photoreceptors. **c, d,** Expression of the nuclear pR8 marker *melt-nlacZ* (green) does not change in *rh6* mutants. In wild-type flies it is expressed with Rh5 (blue) in pR8 and never in Rh6-expressing yR8 cells (red) (**c**). In 5-week-old *rh6* mutants, *melt-nlacZ* is not de-repressed along with Rh5 and remains restricted to pR8 (**d**). **e, f,** Expression of nuclear yR8 marker *wts-nlacZ* does not change in *rh6* mutants. In wild-type flies it is expressed with Rh6 (red) in yR8 and never in Rh5-expressing pR8 cells (blue) (**e**). In *rh6* mutants, *wts-nlacZ* remains in yR8 of 4-week-old flies as Rh5 is de-repressed (**f**).

switch involving mutual transcriptional repression between *warts* (*wts*) and *melted* (*melt*) (Supplementary Fig. 1b). During pupal development, Wts represses *melt* to specify yR8 photoreceptors. In response to an extrinsic signal originating in pR7, *melt* is upregulated in pR8, leading to repression of *wts* transcription and expression of Rh5 (ref. 8). Thus, Melt marks pR8 and Wts marks yR8 cells (Fig. 3c, e). In old *rh6* mutant flies, a *melt* reporter (*melt-nlacZ*) remains restricted to a subset of R8 cells, whereas Rh5 expression expands broadly to cells that do not express *melt-nlacZ* (Fig. 3d and Supplementary Fig. 9d). In addition, we do not observe downregulation of a *wts* reporter (*wts-nlacZ*) in yR8s of old *rh6* mutants, leading to co-expression of *wts* with ectopic Rh5 (Fig. 3f and Supplementary Fig. 9d). Although maintenance of *rh6-lacZ* and *wts-nlacZ* could potentially be due to perdurance of β Gal protein, lack of de-repression of *melt-lacZ* argues that loss of *rh6* function does not affect the identity of yR8 in old flies. Moreover, it shows that *melt* is not involved in Rh5 de-repression. Thus, in *rh6* mutants, the yR8 fate is specified normally and remains stable. This indicates that yR8 produces positive transcriptional regulatory inputs to which the *rh5* promoter can respond and which must be actively

repressed by the presence of Rh6. In contrast to the way pR8 *rh5*-expressing photoreceptor fate is established, these inputs do not depend on extrinsic signals from R7 cells because, as described earlier, the absence of R7s in *sev* mutants does not suppress the *rh6* mutant phenotype.

yR8 cells are not the only photoreceptors expressing Rh6. The larval eye, Bolwig's organ, is composed of about 12 photoreceptors^{21,22}. Four primary photoreceptors express Rh5 whereas the eight secondary photoreceptors express Rh6 (Supplementary Fig. 6a). During metamorphosis, secondary photoreceptors die while the primary photoreceptors downregulate Rh5 and upregulate Rh6 (ref. 23). The newly Rh6-expressing cells form the eyelet, an adult extra-retinal visual organ^{24,25} (Supplementary Fig. 6c). In *rh6* mutants, neither the secondary Bolwig photoreceptors nor the eyelet photoreceptors ever express Rh5 and are thus devoid of any rhodopsin (Supplementary Fig. 6b, d). Therefore, in contrast to the retina, Rh6 is not necessary for exclusion of Rh5 expression in the eyelet, consistent with the view that expression of Rh5 and Rh6 is under distinct control mechanisms in the Bolwig's organ/eyelet and in the adult retina²². This result, together with the absence of Rh5 de-repression in R7s of *rh3* and *rh4* mutants, argues that, in the absence of a rhodopsin signal, de-repression of Rh5 can only occur in yR8 photoreceptors.

Because Rh5 is only de-repressed in yR8s of *rh6* mutants, it is possible that the repressive signal is generated uniquely by Rh6. Therefore, we tested whether the *rh6* mutant phenotype in yR8s could be rescued by rhodopsins other than Rh6. We used *rh6-Gal4* to drive expression of UAS-Rh1, -Rh3, -Rh4 or -Rh6 in *rh6* mutants. In every case, the de-repression was rescued and little or no Rh5 expression was detectable in yR8 photoreceptors (Fig. 4a, b and Supplementary Figs 7b–e and 9e). Expression of UAS-Rh5, as with Rh1, Rh3 and Rh6, also largely blocked de-repression of the *rh5>GFP* reporter in *rh6* mutants (Fig. 4c and Supplementary Fig. 7f–i, 9f), suggesting that a generic *Drosophila* rhodopsin signal is sufficient to maintain exclusion of Rh5 in yR8 cells. Because these transgenes are controlled by the *rh6* promoter, they are expressed only after specification of the yR8 subtype, further arguing that the signal is only required for the maintenance of the exclusion of Rh5, and not for yR8 subtype specification. In addition, negative regulation by Rh5 of its own expression in yR8 could provide an explanation for why the levels of Rh5 expression in yR8 of *rh6* mutants are generally lower than in wild-type pR8 cells.

The requirement for a rhodopsin-dependent signal to maintain repression of *rh5* in yR8s led us to ask whether activation of Rh6 by light is involved in this process. We maintained wild-type flies in complete darkness for more than 2 weeks starting at mid-pupal stages. In these flies, a significant proportion (~12%, Supplementary Table 2) of the Rh6-expressing yR8s also express low levels of Rh5 (Fig. 4d, e and Supplementary Fig. 9g), which is not observed in old wild-type flies reared in the light. Interestingly, this de-repression of Rh5 occurs predominantly in the dorsal retina (Supplementary Table 2 and Supplementary Fig. 9g), indicating an underlying spatial variation in Rh5 de-repression. In contrast, Rh6 is not de-repressed in pR8s of dark-reared flies. Thus, it appears that adult yR8 photoreceptor neurons remain plastic with respect to rhodopsin exclusion and that simply preventing activation of Rh6 by light can evoke Rh5 expression in yR8s. This de-repression of Rh5, however, is substantially weaker than in *rh6* mutants. This could indicate that either activated Rh6 somehow accumulates in the dark and is able to partly repress *rh5* or that Rh6 retains a residual ability to repress *rh5* without being activated by light. These alternatives are consistent with the observation that partial reduction of Rh6 protein through RNA interference can lead to de-repression of *rh5* (Supplementary Fig. 5). Hence, *rh5* repression is sensitive to the level/activity of Rh6.

The role of light and interchangeability of rhodopsins in controlling expression of *rh5* raised the possibility that components of the photo-transduction cascade (reviewed in ref. 5) might play a role in repression of *rh5*. In flies, activated rhodopsin converts the $G_{\alpha q}$ subunit of a

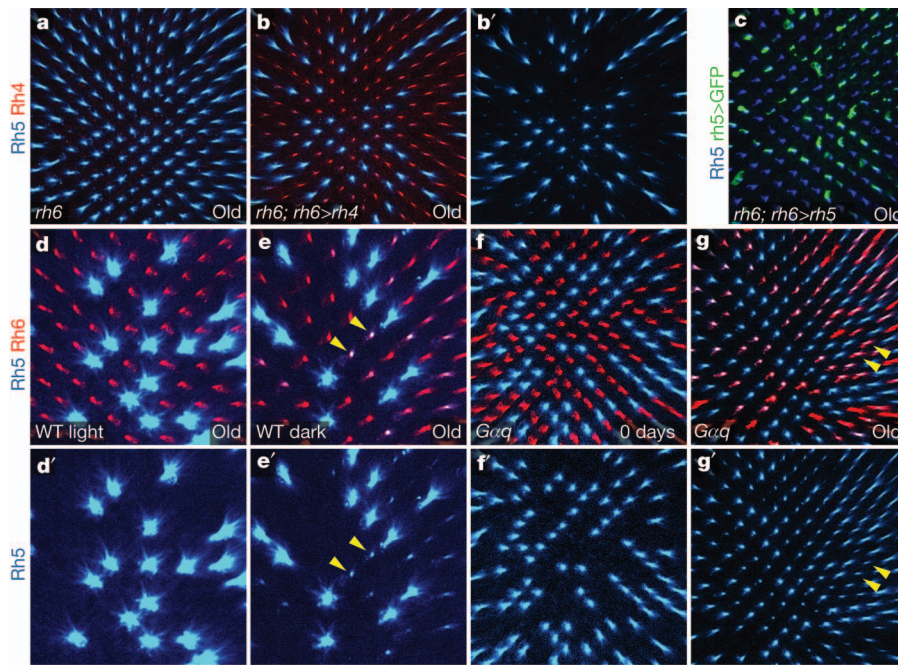


Figure 4 | Part of the phototransduction pathway is required to maintain repression of Rh5. **a, b,** Forced expression of Rh4 (red) in yR8 with *rh6*-Gal4 in *rh6*¹ mutants prevents Rh5 (blue) de-repression (**b**) observed in *rh6*¹ mutant flies (**a**). **c,** Forced expression of Rh5 (blue) in yR8 with *rh6*-Gal4 in *rh6*¹ mutants prevents *rh5*>GFP (green) de-repression observed in *rh6*¹ mutant flies (compare with Fig. 2f and Supplementary Fig. 7f). **d, e,** Dark-reared flies partly de-repress Rh5 in yR8 photoreceptors. In the light, wild-type flies do not

de-repress Rh5 (blue) in Rh6-expressing yR8s (red) (**d**). After 2–3 weeks in complete darkness, a significant number of yR8 cells of wild-type flies express low levels of Rh5 in addition to Rh6 (arrowheads, **e**). **d, e,** Close-ups of dorsal retinas, just dorsal to the equator. **f, g,** *Gαq* is required to maintain repression of Rh5 in yR8. In 2- to 3-week-old (**g**), but not in just eclosed (**f**), *Gαq*¹ mutants, Rh5 (blue) is expressed in yR8 and thus is co-expressed (arrowheads) with Rh6 (red). **b', d'–g',** Rh5 expression alone as in **b, d–g**.

heterotrimeric G-protein to a GTP-bound form which dissociates from the $G_{\beta\gamma}$ dimer and activates phospholipase C encoded by the *norPA* gene. Phospholipase C then catalyses hydrolysis of PIP₂, which leads to the activation of TRPC channels²⁶, inflow of Ca²⁺ and depolarization of the photoreceptors. We asked whether components of this phototransduction cascade mediate the *rh5*-repressive signal. In *Gαq*¹ hypomorphic mutants, Rh5 is expressed normally in young flies but becomes de-repressed in yR8 as the flies age (Fig. 4f, g and Supplementary Fig. 9h), a phenotype similar to that of *rh6* mutants. This results in the co-expression of Rh5 and Rh6 in yR8 cells. However, removal of neither phospholipase C (in *norPA*³⁶ mutants) nor TRPC channels (in *trp1*³⁰²; *trp*³⁰¹ double mutants) leads to de-repression of Rh5 in yR8s of old flies (Supplementary Fig. 8, 9h). The observation that *Gαq*, but not the rest of the phototransduction cascade, is important for the *rh5*-repressive signal indicates a bifurcation of the phototransduction and *rh5*-repression pathways downstream of *Gαq*. Alternatively, *Gαq* might function genetically upstream of Rh6, for example, by stabilizing the Rh6 protein. In either case, Rh6 uses a pathway distinct from phototransduction to repress *rh5*. Importantly, the *Gαq* mutant phenotype and de-repression of Rh5 in dark-raised wild-type flies further support the idea that maintenance of *rh5* repression requires the activity of the Rh6 protein.

Rhodopsins canonically act as sensory receptor proteins. However, Rh1 also has non-visual functions; it is required for the proper formation and maintenance of the rhabdomeres of R1–R6 photoreceptors^{27,28} and has recently been shown to be involved in thermotactic discrimination²⁹. We showed here a new and surprising role for Rh6: it represses transcription of an alternative receptor gene, *rh5*, and thereby maintains the sensory specificity of yR8. This mechanism prevents Rh5/Rh6 co-expression, which would broaden the sensitivity spectrum of yR8 photoreceptors³⁰, limiting the ability of the visual system to discriminate colours. Furthermore, change in the yR8 spectrum could lead to sensory confusion if the downstream neuronal circuits misinterpret the information they receive. The repression of

rh5 by Rh6 also illustrates a so far unappreciated plasticity of yR8 photoreceptor neurons, as revealed by de-repression of Rh5 in wild-type flies reared in darkness. Constant darkness could mimic special environmental conditions, natural for the fly, under which lowered Rh6 activity evokes expression of Rh5 in yR8 photoreceptors to change spectral properties of the eye, or simply to boost the overall light response. Finally, the fact that we found two different *rh6* mutations in laboratory stocks raises a possibility that mutations in the *rh6* gene are also frequent in the natural population.

Repression of *rh5* by Rh6 is reminiscent of the control of olfactory receptor genes in vertebrate olfactory neurons¹⁴, which encode G-protein-coupled receptors similar to Rhodopsins. With rare exception, each olfactory neuron expresses only one allele of one olfactory receptor gene. This exclusion mechanism is not well understood, but requires an active olfactory receptor to generate a feedback signal for repression of other olfactory receptor genes^{10–14}. There, however, the feedback control of exclusion appears to be a common mechanism in all olfactory neurons, in contrast to the fly retina where only *rh5* is regulated by another rhodopsin, and only in the yR8 photoreceptor subtype.

Our findings show that cross-repression of sensory receptors is not unique to vertebrate chemosensory systems, but could be a more widely implemented mechanism by which mature sensory neurons, or other G-protein-coupled receptor-expressing cells, maintain their functional specificity. The relative simplicity of yR8 photoreceptors as a system should allow us to uncover the molecular details by which a G-protein-coupled receptor can exclude expression of other seven-transmembrane receptors in the same cell.

METHODS SUMMARY

Flies were raised on standard corn meal/molasses/agar medium at room temperature (24 °C) in ambient laboratory light except for RNAi experiments (at 29 °C) and dark isolation experiments (in complete darkness). Dissected adult retinas were stained wholemount with specific primary antibodies and then with Alexa Fluor-conjugated secondary antibodies (Molecular Probes). Larval eyes were stained as in ref. 22. *In situ* hybridization on cryo-sectioned adult retinas was

performed with DIG-labelled antisense probe transcribed from *rh5* 3' untranslated region (UTR) as described in ref. 7. Samples were imaged using Leica TCS SP2 and SP5 confocal microscopes. Images were processed and counts performed using Leica Confocal Software, Adobe Photoshop and Fiji software. For real-time PCR, RNA was purified from 20 flies per sample and complementary DNA (cDNA) amplified using SYBR-Green PCR Mix (Stratagene).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 October 2010; accepted 15 August 2011.

Published online 9 October 2011.

1. Mombaerts, P. Axonal wiring in the mouse olfactory system. *Annu. Rev. Cell Dev. Biol.* **22**, 713–737 (2006).
2. Komiyama, T. & Luo, L. Development of wiring specificity in the olfactory system. *Curr. Opin. Neurobiol.* **16**, 67–73 (2006).
3. Morey, M. *et al.* Coordinate control of synaptic-layer specificity and rhodopsins in photoreceptor neurons. *Nature* **456**, 795–799 (2008).
4. Rister, J. & Desplan, C. The retinal mosaics of opsin expression in invertebrates and vertebrates. *Dev. Neurobiol.* doi:10.1002/dneu.20905 (2011).
5. Wang, T. & Montell, C. Phototransduction and retinal degeneration in *Drosophila*. *Pflügers Arch.* **454**, 821–847 (2007).
6. Franceschini, N., Kirschfeld, K. & Minke, B. Fluorescence of photoreceptor cells observed *in vivo*. *Science* **213**, 1264–1267 (1981).
7. Mazzoni, E. O. *et al.* Iroquois complex genes induce co-expression of rhodopsins in *Drosophila*. *PLoS Biol.* **6**, e97 (2008).
8. Mikeladze-Dvali, T. *et al.* The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors. *Cell* **122**, 775–787 (2005).
9. Wernet, M. F. *et al.* Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174–180 (2006).
10. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
11. Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801–815 (2004).
12. Lewcock, J. W. & Reed, R. A. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl Acad. Sci. USA* **101**, 1069–1074 (2004).
13. Feinstein, P., Bozza, T., Rodriguez, I., Vassalli, A. & Mombaerts, P. Axon guidance of mouse olfactory sensory neurons by odorant receptors and the $\beta 2$ adrenergic receptor. *Cell* **117**, 833–846 (2004).
14. Fuss, S. H. & Ray, A. Mechanisms of odorant receptor gene choice in *Drosophila* and vertebrates. *Mol. Cell. Neurosci.* **41**, 101–112 (2009).
15. Cook, T., Pichaud, F., Sonnevill, R., Papatsenko, D. & Desplan, C. Distinction between color photoreceptor cell fates is controlled by Prospero in *Drosophila*. *Dev. Cell* **4**, 853–864 (2003).
16. Harris, W. A., Stark, W. S. & Walker, J. A. Genetic dissection of the photoreceptor system in the compound eye of *Drosophila melanogaster*. *J. Physiol. (Lond.)* **256**, 415–439 (1976).
17. Chou, W. H. *et al.* Patterning of the R7 and R8 photoreceptor cells of *Drosophila*: evidence for induced and default cell-fate specification. *Development* **126**, 607–616 (1999).
18. Papatsenko, D., Sheng, G. & Desplan, C. A new rhodopsin in R8 photoreceptors of *Drosophila*: evidence for coordinate expression with Rh3 in R7 cells. *Development* **124**, 1665–1673 (1997).
19. Chou, W. H. *et al.* Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17**, 1101–1115 (1996).
20. Tahayato, A. *et al.* Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the *Drosophila* retina. *Dev. Cell* **5**, 391–402 (2003).
21. Green, P., Hartenstein, A. Y. & Hartenstein, V. The embryonic development of the *Drosophila* visual system. *Cell Tissue Res.* **273**, 583–598 (1993).
22. Sprecher, S. G., Pichaud, F. & Desplan, C. Adult and larval photoreceptors use different mechanisms to specify the same Rhodopsin fates. *Genes Dev.* **21**, 2182–2195 (2007).
23. Sprecher, S. G. & Desplan, C. Switch of rhodopsin expression in terminally differentiated *Drosophila* sensory neurons. *Nature* **454**, 533–537 (2008).
24. Hofbauer, A. & Buchner, E. Does *Drosophila* have seven eyes? *Naturwissenschaften* **76**, 335–336 (1989).
25. Yasuyama, K. & Meinertzhagen, I. A. Extraretinal photoreceptors at the compound eye's posterior margin in *Drosophila melanogaster*. *J. Comp. Neurol.* **412**, 193–202 (1999).
26. Huang, J. *et al.* Activation of TRP channels by protons and phosphoinositide depletion in *Drosophila* photoreceptors. *Curr. Biol.* **20**, 189–197 (2010).
27. O'Tousa, J. E., Leonard, D. S. & Pak, W. L. Morphological defects in oraJK84 photoreceptors caused by mutation in R1–6 opsin gene of *Drosophila*. *J. Neurogenet.* **6**, 41–52 (1989).
28. Kumar, J. P. & Ready, D. F. Rhodopsin plays an essential structural role in *Drosophila* photoreceptor development. *Development* **121**, 4359–4370 (1995).
29. Shen, W. L. *et al.* Function of rhodopsin in temperature discrimination in *Drosophila*. *Science* **331**, 1333–1336 (2011).
30. Arikawa, K., Mizuno, S., Kinoshita, M. & Stavenga, D. G. Coexpression of two visual pigments in a photoreceptor causes an abnormally broad spectral sensitivity in the eye of the butterfly *Papilio xuthus*. *J. Neurosci.* **23**, 4527–4532 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Blau, B. Collins, M. Cols, T. Erclik, S.H. Fuss, D. Jukam, J.P. Kumar, E. Laufer, H.-S. Li, B. Minke, C. Montell, F. Pichaud, J. Rister and A. Tomlinson for suggestions and comments on the manuscript, V. Douard for help with qRT-PCR, J. Goodness for help identifying *rh6^{ts}* allele, S.G. Britt, P.J. Dolph, P.R. Hiesinger, F. Pichaud, N. Pinal, D.F. Ready, C.S. Zuker, and the Bloomington *Drosophila* Stock Center for flies or antibodies. This work was funded by the National Institutes of Health R01EY13012 to C.D. and F32EY016309 to D.V.

Author Contributions D.V., E.O.M. and C.D. conceived the experiments; D.V. and E.O.M. designed and performed experiments in adult flies; S.G.S. designed and performed experiments in larvae; K.B. performed RNAi experiments; R.J.J., P.L., N.V. and A.C. contributed reagents; D.V. and C.D. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.D. (cd38@nyu.edu).

METHODS

Flies were raised on standard cornmeal/molasses/agar medium at room temperature (24 °C) in ambient laboratory light unless otherwise noted. RNAi experiments were performed at 29 °C. For dark isolation experiments, flies were reared in a lightproof box, and for ageing transferred between vials in complete darkness starting at mid-pupal stages (before rhodopsin expression³¹).

Drosophila strains. For wild-type controls we used $y^1 w^{67}$; Sp/CyO ; wt^{isoB} flies, where 'isoB' represents an isogenized wild-type third chromosome.

We used the following *rh6* alleles.

The *rh6¹* allele¹⁵ is found in many commonly used laboratory fly strains. The existence of this mutation in common stocks was originally pointed out to us by S. Britt. This mutation is present on some TM6B balancer chromosomes and in the reference fly strain sequenced for the published fly genome¹⁵ (BDGP release 5.29). The mutation replaces 21 bases (lower case in TGACCATCATCTTCTctac tggcacatcatgaaggTATGACATTCGTTA) at the end of the first exon with two As, removing a splice donor site and introducing a stop codon immediately afterwards. This results in the truncation of the open reading frame within the fifth transmembrane domain of the presumptive protein. The original allele was backcrossed into wt^{isoB} background (see above) four times.

We identified *rh6⁵* as a mutation in a stock from the Bloomington *Drosophila* Stock Center (Stock 1385, named genotype $z^{v77h} w^{67c23}$) which mapped to the third chromosome. Sequencing of *rh6* locus revealed a 58-bp deletion upstream of the *rh6* transcription start site, which removes sequence AGCGGCAATCGAAAGCC CAATTGCAACGGTTAGCTTTGGATTGGCCAAAGTGCCGGCTA within the *rh6* promoter. We named this mutation after the singer Frank Sinatra, for his nickname "Ol' blue eyes", because eyes of old *rh6⁵* mutant flies broadly express the blue-sensitive rhodopsin, Rh5.

The deficiency used in this study that covers *rh6* gene, *Df(3R)Exel6174*, was generated by Exelixis and spans 3R:11154443–11154444 ... 11363188 (ref. 32).

To generate flies with a *rh6* genomic rescue fragment, *C(rh6⁺)*, the *rh6* sequence was PCR-amplified from genomic DNA of $y^1 w^{67}$; Sp/CyO ; wt^{isoB} flies with dv173 (ACAAGCTTACCTACAAGAGCACCAGTCC) and dv174 (ACGAATTCA CCTCGGCCTGAACACCTAC) primers to produce a 2,575-bp genomic fragment (ACCTACAAGAGCACCAGTCC ... GTAGGTGTCAGGCCGAGGT) with HindIII and EcoRI flanking sites. PCR product was ligated into HindIII–EcoRI sites of pBS-loxP-w-lox2272 vector³³. Cre-recombinase-mediated integration was used to insert this construct into lox landing site A11 (on the second chromosome, S. Small, personal communication). A single integration occurred without replacement of y^1 marker of the landing site. Successful transformation was confirmed with antibody stain for Rh6 protein in wholemount retinas: normal Rh6 expression was detected in *rh6* mutant background.

UAS-rh6RNAi (transformant 102152) was obtained from the Vienna *Drosophila* RNAi Center³⁴.

We generated the following additional mutants.

rh3¹ mutant (a nucleotide change C278T resulting in Q46* truncation) was obtained by TILLING (Seattle TILLING Project)³⁵. The mutation was backcrossed into wild-type background four times (confirmed by genomic PCR and by stain of wholemount retinas with anti-Rh3 antibody).

rh4¹ mutant (a nucleotide change T727A resulting in Y203* truncation between fourth and fifth transmembrane domains) was obtained by TILLING (Seattle TILLING Project)³⁵. The mutation was back crossed into wild-type background four times (confirmed by genomic PCR and by stain of wholemount retinas with anti-Rh4 antibody).

We generated the following transgenic lines.

rh5>GFP flies carry two transgenes recombined on the second chromosome: *rh5-lexA* and *lexAop-GFP*.

rh5-lexA. *lexA* (from pBS-lexA SV40 3' UTR³⁶) was cloned into pBS-LoxP-white-Lox2272 (ref. 33) and named LexA-Lox, a 740-bp fragment of *rh5* promoter which ends 23 bases upstream of ATG (TCGGAAAATGTCGTGCAAGTGTTTC ... AATGTCGACCTGCAAAGGAAACTA; fly genome: 12007686 ... 12008425) was PCR amplified from genomic DNA using oBJ109 (TCGGAAAATGT CGTGCAGTG) and oBJ140 (TAGTTTCCTTTGCAGGTCGAC) and cloned into the PCRII-TOPO vector (Invitrogen). The *rh5* promoter was cut with ClaI, blunted and subcloned into the LexA-Lox, which was cut with SpeI and blunted. Cre-recombinase-mediated cassette exchange was used to insert this construct onto the second chromosome³³.

lexAop-GFP. GFP with SV40 3' UTR was PCR amplified from the pIRES2-eGFP vector (Clontech) with the primers oBJ78 (TAATACTAGTATGGTGA GCAAGGGCGAGGAG) and oBJ79 (GTCAGGATCCACCACAAGTAGAATG CAGTG) and cloned into the PCRII-TOPO vector (Invitrogen). The GFP-SV40 3' UTR was subcloned into the pLOT vector (containing *lexAop*)³⁶ using the EcoRI site.

UAS-Rh1. EcoRI–KpnI fragment of *rh1* cDNA (containing sequence spanned by GGCAGGTTTCCAACGACCAATCGC ... AAGGACAAAAAAACTCA AC + 15A) from rh1-pFLC-1 plasmid (*Drosophila* Genomics Resource Center clone RH01460 (ref. 37)) was ligated into EcoRI–KpnI sites of pUASTattB vector³⁸ to produce pDV131 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing sites *attP-51D*, *attP-58A* and *attP40* (refs 38, 39). w^+ and *3xP3-RFP* markers of *attP-51D* and *attP-58A* landing sites were removed through lox-mediated recombination by crossing in Cre recombinase transgene³⁸.

UAS-Rh3. EcoRI–XhoI fragment of *rh3* cDNA (containing sequence spanned by CAGACCGGAGCATGGAGTCCGGTA ... AATATAGTAAATTACAGC AAGCT + 19A) from rh3-pOT2 (*Drosophila* Genomics Resource Center clone GH02505 (ref. 37)) was ligated into EcoRI–XhoI sites of pUASTattB vector³⁸ to produce pDV133 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing sites *attP-51D*, *attP-58A* and *attP40* (refs 38, 39). w^+ and *3xP3-RFP* markers of *attP-51D* and *attP-58A* landing sites were removed through lox-mediated recombination by crossing in Cre recombinase transgene³⁸.

UAS-Rh4. Cloned EcoRI–KpnI fragment of *rh4* cDNA from rh4-pFLC-1 (*Drosophila* Genomics Resource Center clone RH33063 (ref. 37)) was ligated into pUASTattB vector³⁸. To correct a frameshift in the sequence, EcoRI–BglII fragment was replaced with cDNA fragment that had a longer 5' UTR. The resulting pDV134 plasmid contained *rh4* cDNA sequence spanned by CAGAGCGAAAC GGGTAGCGGT ... AACTTATTGCAAACGAAGTAG + 16A. ϕ C31-mediated integration was used to insert this construct into second chromosome landing sites *attP-51D* and *attP40* (refs 38, 39). w^+ and *3xP3-RFP* markers of *attP-51D* landing site were removed through lox-mediated recombination by crossing in Cre recombinase transgene³⁸.

UAS-Rh5. EcoRI–XhoI fragment of *rh5* cDNA (containing sequence spanned by CGGAGGCCAGAATGTCGACCT ... TACAAACCAAAAAAAGTTGGCA TT + 78A) from rh5-pOT2 (*Drosophila* Genomics Resource Center clone GH28578 (ref. 37)) was ligated into EcoRI–XhoI sites of pUASTattB vector³⁸ to produce pDV135 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing site *attP40* (ref. 39).

UAS-Rh6. It has proved difficult to generate a *UAS-Rh6* cDNA construct expressing high levels of Rh6. Therefore, we cloned a PCR-amplified genomic (with introns) fragment of *rh6* gene downstream of transcriptional start site (containing sequence spanned by CAGGCATTGCCGCGAGTTCGCGT ... ACAG CAATTGATACAAAATC) into EcoRI–KpnI sites of pUASTattB vector³⁸ to produce pDV160 plasmid. ϕ C31-mediated integration was used to insert this construct into second chromosome landing site *attP40* (ref. 39).

We used the following other strains.

G_{2q}^1 (ref. 40), *norpA³⁶* (ref. 41), *rh5²* (ref. 42), *sev¹⁴* (ref. 43), *trpl³⁰²*; *trp³⁰¹* (ref. 44), *melt-nlacZ⁸*, *rh6-Gal4* (ref. 20), *rh6-lacZ⁴⁵* and *wts-nlacZ^{46,47}*.

Antibodies. Antibodies and dilutions used were as follows: mouse anti-Rh1 (1:10) (DSHB, clone 4C5); mouse anti-Rh3 (1:10) and mouse anti-Rh5 (1:100) (gifts from S. Britt); rabbit anti-Rh4 (1:100) (gift from C. Zuker); rabbit anti-Rh6 (1:2,000) (ref. 20); goat anti- β Gal (1:5,000) (Biogenesis); mouse anti- β Gal (1:500) (Promega); rat anti-Elav (1:40) (DSHB, clone Rat-Elav-7E8A10); sheep anti-GFP (1:500) (AbD Serotec); rabbit anti-GFP (1:800) (Biogenesis). Secondary antibodies raised in donkey and goat were Alexa-Fluor-conjugated (Alexa Fluor 488 at 1:1,000, Alexa Fluor 555 at 1:750, Alexa Fluor 647 at 1:500) (Molecular Probes). Alexa Fluor 488-conjugated phalloidin was used to visualize rhabdomeres (1:100, Molecular Probes).

Stains. Adult retinas were dissected out in phosphate buffered saline (PBS), fixed for 15 min with 4% formaldehyde at room temperature, washed three times in PBS, and incubated with the primary antibodies diluted in Block (PBS, 0.1% Triton-X-100, 2% horse serum) overnight at 4 °C. After two rinses and two 1-h washes with PBT (PBS, 0.3% Triton-X-100), the retinas were incubated overnight at 4 °C with secondary antibodies diluted in Block. Retinas were rinsed twice and after two 1-h washes with PBT, were mounted in SlowFade Gold (Invitrogen). Antibody staining for larval eye was performed as described in ref. 22. *In situ* hybridization for cryo-sectioned adult retinas was performed as described in ref. 7 with DIG-labelled antisense probe transcribed from cloned rh5 3' UTR region (bp 900–1411). Samples were imaged using Leica TCS SP2 and SP5 confocal microscopes. Images were processed using Leica Confocal Software (LCS), Adobe Photoshop and Fiji software.

Counting. Optical sections were photographed approximately 10 μ m distal to R8 nuclei in the centre of the retina. The portion of the image of the retina section containing R8 rhabdomeres was defined as area populated with Rh5-positive cells. The number of Rh5-expressing R8s and the total number of R8s (represented by ommatidia visualized with phalloidin) in this area were counted using Fiji software with Cell Counter plug in.

RNA analysis. RNA was purified from each sample of about 20 flies with TRIzol (Invitrogen), RNeasy mini columns (Qiagen) and treated with DNaseI (Qiagen). Three micrograms of total RNA was reverse transcribed with oligo(dT)20 and SuperScript III Reverse Transcriptase (Invitrogen). The cDNA was amplified in duplicate reactions using SYBR-Green PCR Mix (Stratagene) by real-time PCR. Primers used are listed in Supplementary Table 3. Target gene levels were normalized to levels of *rp49* mRNA⁴⁸ and expressed relative to levels in 0-day-old wild-type flies. At least three independent replicates were averaged for each experimental condition.

31. Earl, J. B. & Britt, S. G. Expression of *Drosophila* rhodopsins during photoreceptor cell differentiation: insights into R7 and R8 cell subtype commitment. *Gene Expr. Patterns* **6**, 687–694 (2006).
32. Parks, A. L. *et al.* Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nature Genet.* **36**, 288–292 (2004).
33. Oberstein, A., Pare, A., Kaplan, L. & Small, S. Site-specific transgenesis by Cre-mediated recombination in *Drosophila*. *Nature Methods* **2**, 583–585 (2005).
34. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
35. Till, B. J. *et al.* High-throughput TILLING for functional genomics. *Methods Mol. Biol.* **236**, 205–220 (2003).
36. Lai, S. L. & Lee, T. Genetic mosaic with dual binary transcriptional systems in *Drosophila*. *Nature Neurosci.* **9**, 703–709 (2006).
37. Stapleton, M. *et al.* The *Drosophila* gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res.* **12**, 1294–1300 (2002).
38. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
39. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. & Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature Genet.* **40**, 476–483 (2008).
40. Scott, K., Becker, A., Sun, Y., Hardy, R. & Zuker, C. Gq alpha protein function in vivo: genetic dissection of its role in photoreceptor cell physiology. *Neuron* **15**, 919–927 (1995).
41. Bloomquist, B. T. *et al.* Isolation of a putative phospholipase C gene of *Drosophila*, *norpA*, and its role in phototransduction. *Cell* **54**, 723–733 (1988).
42. Yamaguchi, S., Wolf, R., Desplan, C. & Heisenberg, M. Motion vision is independent of color in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105**, 4910–4915 (2008).
43. Gerresheim, F. Isolation of *Drosophila melanogaster* mutants with a wavelength-specific alteration in their phototactic response. *Behav. Genet.* **18**, 227–246 (1988).
44. Niemeyer, B. A., Suzuki, E., Scott, K., Jalink, K. & Zuker, C. S. The *Drosophila* light-activated conductance is composed of the two channels TRP and TRPL. *Cell* **85**, 651–659 (1996).
45. Papatsenko, D., Nazina, A. & Desplan, C. A conserved regulatory element present in all *Drosophila* rhodopsin genes mediates Pax6 functions and participates in the fine-tuning of cell-specific expression. *Mech. Dev.* **101**, 143–153 (2001).
46. Justice, R. W., Zilian, O., Woods, D. F., Noll, M. & Bryant, P. J. The *Drosophila* tumor suppressor gene *warts* encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes Dev.* **9**, 534–546 (1995).
47. Xu, T., Wang, W., Zhang, S., Stewart, R. A. & Yu, W. Identifying tumor suppressors in genetic mosaics: the *Drosophila* *lats* gene encodes a putative protein kinase. *Development* **121**, 1053–1063 (1995).
48. O’Keefe, L. V. *et al.* *Drosophila* orthologue of WWOX, the chromosomal fragile site FRA16D tumour suppressor gene, functions in aerobic metabolism and regulates reactive oxygen species. *Hum. Mol. Genet.* **20**, 497–509 (2011).

Solutions for a cultivated planet

Jonathan A. Foley¹, Navin Ramankutty², Kate A. Brauman¹, Emily S. Cassidy¹, James S. Gerber¹, Matt Johnston¹, Nathaniel D. Mueller¹, Christine O'Connell¹, Deepak K. Ray¹, Paul C. West¹, Christian Balzer³, Elena M. Bennett⁴, Stephen R. Carpenter⁵, Jason Hill^{1,6}, Chad Monfreda⁷, Stephen Polasky^{1,8}, Johan Rockström⁹, John Sheehan¹, Stefan Siebert¹⁰, David Tilman^{1,11} & David P. M. Zaks¹²

Increasing population and consumption are placing unprecedented demands on agriculture and natural resources. Today, approximately a billion people are chronically malnourished while our agricultural systems are concurrently degrading land, water, biodiversity and climate on a global scale. To meet the world's future food security and sustainability needs, food production must grow substantially while, at the same time, agriculture's environmental footprint must shrink dramatically. Here we analyse solutions to this dilemma, showing that tremendous progress could be made by halting agricultural expansion, closing 'yield gaps' on underperforming lands, increasing cropping efficiency, shifting diets and reducing waste. Together, these strategies could double food production while greatly reducing the environmental impacts of agriculture.

Contemporary agriculture faces enormous challenges^{1–3}. Even with recent productivity gains, roughly one in seven people lack access to food or are chronically malnourished, stemming from continued poverty and mounting food prices^{4,5}. Unfortunately, the situation may worsen as food prices experience shocks from market speculation, bioenergy crop expansion and climatic disturbances^{6,7}. Even if we solve these food access challenges, much more crop production will probably be needed to guarantee future food security. Recent studies suggest that production would need to roughly double to keep pace with projected demands from population growth, dietary changes (especially meat consumption), and increasing bioenergy use^{1–4,8,9}, unless there are dramatic changes in agricultural consumption patterns.

Compounding this challenge, agriculture must also address tremendous environmental concerns. Agriculture is now a dominant force behind many environmental threats, including climate change, biodiversity loss and degradation of land and freshwater^{10–12}. In fact, agriculture is a major force driving the environment beyond the “planetary boundaries” of ref. 13.

Looking forward, we face one of the greatest challenges of the twenty-first century: meeting society's growing food needs while simultaneously reducing agriculture's environmental harm. Here we consider several promising solutions to this grand challenge. Using new geospatial data and models, we evaluate how new approaches to agriculture could benefit both food production and environmental sustainability. Our analysis focuses on the agronomic and environmental aspects of these challenges, and leaves a richer discussion of associated social, economic and cultural issues to future work.

The state of global agriculture

Until recently, the scientific community could not measure, monitor and analyse the agriculture–food–environment system's complex linkages at the global scale. Today, however, we have new data that characterize worldwide patterns and trends in agriculture and the environment^{14–17}.

Agricultural extent

According to the Food and Agriculture Organization (FAO) of the United Nations, croplands cover 1.53 billion hectares (about 12% of Earth's ice-free land), while pastures cover another 3.38 billion hectares (about 26% of Earth's ice-free land) (Supplementary Fig. 1). Altogether, agriculture occupies about 38% of Earth's terrestrial surface—the largest use of land on the planet^{14,18}. These areas comprise the land best suited for farming¹⁹: much of the rest is covered by deserts, mountains, tundra, cities, ecological reserves and other lands unsuitable for agriculture²⁰.

Between 1985 and 2005 the world's croplands and pastures expanded by 154 million hectares (about 3%). But this slow net increase includes significant expansion in some areas (the tropics), as well as little change or a decrease in others (the temperate zone¹⁸; Supplementary Table 1). The result is a net redistribution of agricultural land towards the tropics, with implications for food production, food security and the environment.

Crop yields

Global crop production has increased substantially in recent decades. Studies of common crop groups (including cereals, oilseeds, fruits and vegetables) suggest that crop production increased by 47% between 1985 and 2005 (ref. 18). However, considering all 174 crops tracked by the UN FAO and ref. 15, we find global crop production increased by only 28% during that time¹⁸.

This 28% gain in production occurred as cropland area increased by only 2.4%, suggesting a 25% increase in yield. However, cropland area that was harvested increased by about 7% between 1985 and 2005—nearly three times the change in cropland area, owing to increased multiple cropping, fewer crop failures, and less land left fallow. Accounting for the increase in harvested land, average global crop yields increased by only 20% between 1985 and 2005, substantially less than the often-cited 47% production increase for selected crop groups. (Using the same methods as for the 20% result, we note that yields increased by 56% between 1965 and 1985, indicating that yields are now rising less quickly than before.)

¹Institute on the Environment (IonE), University of Minnesota, 1954 Buford Avenue, Saint Paul, Minnesota 55108, USA. ²Department of Geography and Global Environmental and Climate Change Centre, McGill University, 805 Sherbrooke Street, West Montreal, Quebec H3A 2K6, Canada. ³Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA. ⁴School of Environment and Department of Natural Resource Sciences, McGill University, 111 Lakeshore Road, Ste Anne de Bellevue, Quebec H9X 3V9, Canada. ⁵Center for Limnology, University of Wisconsin, 680 North Park Street, Madison, Wisconsin 53706, USA. ⁶Department of Bioproducts and Biosystems Engineering, University of Minnesota, 2004 Folwell Avenue, Minnesota 55108, USA. ⁷Consortium for Science, Policy and Outcomes (CSPO), Arizona State University, 1120 S Cady Mall, Tempe, Arizona 85287, USA. ⁸Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, Minnesota 55108, USA. ⁹Stockholm Resilience Centre, Stockholm University, SE-106 91, Stockholm, Sweden. ¹⁰Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5, D53115, Bonn, Germany. ¹¹Department of Ecology, Evolution & Behavior, University of Minnesota, 1987 Upper Buford Circle, Minnesota 55108, USA. ¹²Center for Sustainability and the Global Environment (SAGE), University of Wisconsin, 1710 University Avenue, Madison, Wisconsin 53726, USA.

Aggregate measures of production can mask trends in individual crops or crop groups (Supplementary Fig. 2a). For example, cereal crops decreased in harvested area by 3.6% between 1985 and 2005, yet their total production increased by 29%, reflecting a 34% increase in yields per hectare. Oil crops, on the other hand, showed large increases in both harvested area (43%) and yield (57%), resulting in a 125% increase in total production¹⁸. While most crops increased production between 1985 and 2005, fodder crops did not: on average, they saw an 18% production drop as a 26% loss in harvested area overrode an 11% increase in yields.

Using geospatial data¹⁵, we can examine how yield patterns have changed for key commodities (for example, maize in Supplementary Fig. 2b). These geographic patterns show us where productivity gains have been successful, where they have not, and where further opportunities for improvement lie.

Crop use and allocation

The allocation of crops to nonfood uses, including animal feed, seed, bioenergy and other industrial products, affects the amount of food available to the world. Globally, only 62% of crop production (on a mass basis) is allocated to human food, versus 35% to animal feed (which produces human food indirectly, and much less efficiently, as meat and dairy products) and 3% for bioenergy, seed and other industrial products.

A striking disparity exists between regions that primarily grow crops for direct human consumption and those that produce crops for other uses (Fig. 1). North America and Europe devote only about 40% of their croplands to direct food production, whereas Africa and Asia allocate typically over 80% of their cropland to food crops. Extremes range from the Upper Midwestern USA (less than 25%) to South Asia (over 90%).

As we face the twin challenges of feeding a growing world while charting a more environmentally sustainable path, the amount of land (and other resources) devoted to animal-based agriculture merits critical evaluation. For example, adding croplands devoted to animal feed (about 350 million hectares) to pasture and grazing lands (3.38 billion hectares), we find the land devoted to raising animals totals 3.73 billion hectares—an astonishing ~75% of the world's agricultural land. We further note that meat and dairy production can either add to or subtract

from the world's food supply. Grazing systems, especially on pastures unsuitable for other food production, and mixed crop–livestock systems can add calories and protein to the world and improve economic conditions and food security in many regions. However, using highly productive croplands to produce animal feed, no matter how efficiently, represents a net drain on the world's potential food supply.

Global environmental impacts of agriculture

The environmental impacts of agriculture include those caused by expansion (when croplands and pastures extend into new areas, replacing natural ecosystems) and those caused by intensification (when existing lands are managed to be more productive, often through the use of irrigation, fertilizers, biocides and mechanization). Below, we use new data and models^{17,21,22} to examine both.

Agricultural expansion has had tremendous impacts on habitats, biodiversity, carbon storage and soil conditions^{10,11,23,24}. In fact, worldwide agriculture has already cleared or converted 70% of the grassland, 50% of the savanna, 45% of the temperate deciduous forest, and 27% of the tropical forest biome^{14,25}.

Today, agriculture is mainly expanding in the tropics, where it is estimated that about 80% of new croplands are replacing forests²⁶. This expansion is worrisome, given that tropical forests are rich reservoirs of biodiversity and key ecosystem services²⁷. Clearing tropical forests is also a major source of greenhouse gas emissions and is estimated to release around 1.1×10^{15} grams of carbon per year, or about 12% of total anthropogenic CO₂ emissions²⁸. Slowing or halting expansion of agriculture in the tropics—which accounts for 98% of total CO₂ emissions from land clearing²⁹—will reduce carbon emissions as well as losses of biodiversity and ecosystem services²⁷.

Agricultural intensification has dramatically increased in recent decades, outstripping rates of agricultural expansion, and has been responsible for most of the yield increases of the past few decades. In the past 50 years, the world's irrigated cropland area roughly doubled^{18,30,31}, while global fertilizer use increased by 500% (over 800% for nitrogen alone)^{18,32,33}. Intensification has also caused water degradation, increased energy use, and widespread pollution^{32,34,35}.

Of particular concern is that some 70% of global freshwater withdrawals (80–90% of consumptive uses) are devoted to irrigation^{36,37}. Furthermore, rain-fed agriculture is the world's largest user of water^{13,38}. In addition, fertilizer use, manure application, and leguminous crops (which fix nitrogen in the soil) have dramatically disrupted global nitrogen and phosphorus cycles^{39–41}, with associated impacts on water quality, aquatic ecosystems and marine fisheries^{35,42}.

Both agricultural expansion and intensification are also major contributors to climate change. Agriculture is responsible for 30–35% of global greenhouse gas emissions, largely from tropical deforestation, methane emissions from livestock and rice cultivation, and nitrous oxide emissions from fertilized soils^{29,43–46}.

We can draw important conclusions from these trends. First, the expansion of agriculture in the tropics is reducing biodiversity, increasing greenhouse gas emissions, and depleting critical ecosystem services. Yet this expansion has done relatively little to add to global food supplies; most production gains have been achieved through intensification. Second, the costs and benefits of agricultural intensification vary greatly, often depending on geographic conditions and agronomic practices. This suggests that some forms (and locations) of intensification are better than others at balancing food production and environmental protection^{11,47}.

Enhancing food production and sustainability

Until recently, most agricultural paradigms have focused on improving production, often to the detriment of the environment^{10,11,47}. Likewise, many environmental conservation strategies have not sought to improve food production. However, to achieve global food security and environmental sustainability, agricultural systems must be transformed to address both challenges (Fig. 2).

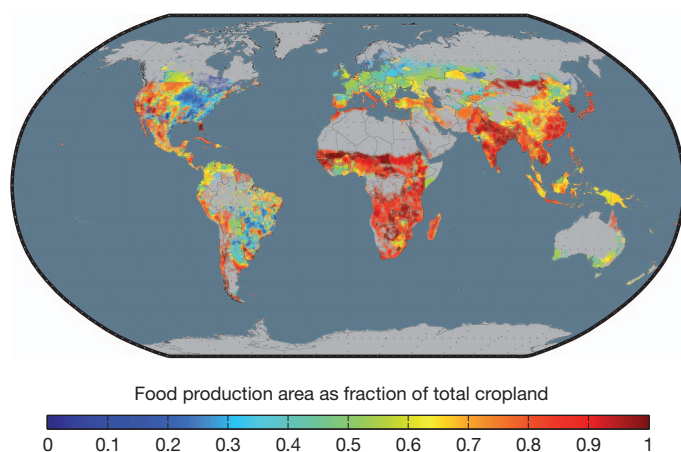


Figure 1 | Allocation of cropland area to different uses in 2000. Here we show the fraction of the world's total cropland that is dedicated to growing food crops (crops that are directly consumed by people) versus all other crop uses, including animal feed, fibre, bioenergy crops and other products. Averaged across the globe, 62% of total crop production (on a mass basis) is allocated to human food, 35% for animal feed (which produces human food indirectly, and less efficiently, as meat and dairy products) and 3% for bioenergy crops, seed, and other industrial products. There are striking disparities between regions that primarily grow crops for human consumption (such as Africa, South Asia, East Asia), and those that mainly produce crops for other uses (such as North America, Europe, Australia). Food production and allocation data were obtained from FAOSTAT¹⁸, and were then applied to the spatial cropland maps of refs 14 and 15. All data are for a seven-year period centred on 2000.

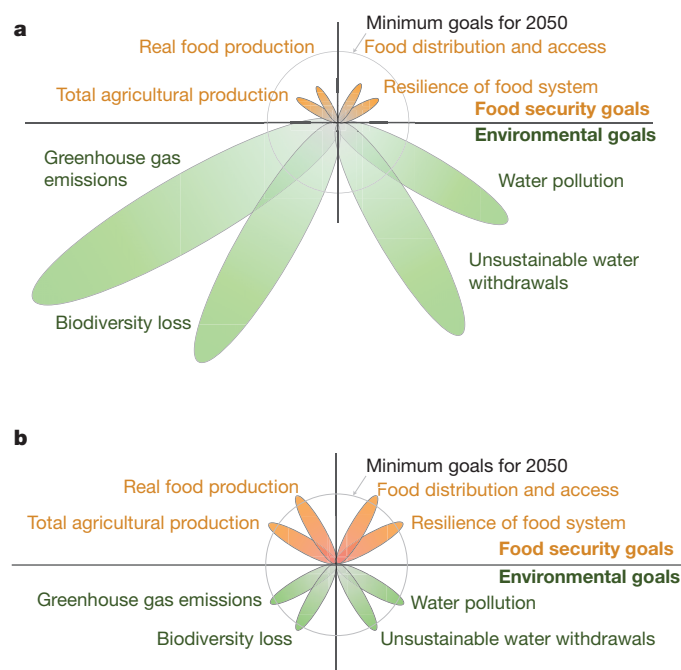


Figure 2 | Meeting goals for food security and environmental sustainability by 2050. Here we qualitatively illustrate a subset of the goals agriculture must meet in the coming decades. At the top, we outline four key food security goals: increasing total agricultural production, increasing the supply of food (recognizing that agricultural yields are not always equivalent to food), improving the distribution of and access to food, and increasing the resilience of the whole food system. At the bottom, we illustrate four key environmental goals agriculture must also meet: reducing greenhouse gas emissions from agriculture and land use, reducing biodiversity loss, phasing out unsustainable water withdrawals, and curtailing air and water pollution from agriculture. Panel **a** sketches out a qualitative assessment of how current agricultural systems may be measured against these criteria compared to goals set for 2050. Panel **b** illustrates a hypothetical situation in which we meet all of these goals by 2050.

First, the transformation of agriculture must deliver sufficient food and nutrition to the world. To meet the projected demands of population growth and increasing consumption, we must roughly double food supplies in the next few decades^{1–3}. We must also improve distribution and access, which will require further changes in the food system.

The transformation of agriculture should also (1) cut greenhouse gas emissions from land use and farming by at least 80% (ref. 48); (2) reduce biodiversity and habitat losses; (3) reduce unsustainable water withdrawals, especially where water has competing demands; and (4) phase out water pollution from agricultural chemicals. Other environmental issues must also be addressed, but these four undergird the relationship between agriculture and the environment and should be addressed as necessary first steps.

An influential series of recent reports has suggested possible solutions to our interwoven food security and environmental challenges^{1,2,6}. Below, we consider the potential strengths and weaknesses of four proposed strategies.

Stop expanding agriculture

The expansion of agriculture into sensitive ecosystems has far-reaching effects on biodiversity, carbon storage and important environmental services^{10,11,33}. This is particularly true when tropical forests are cleared for agriculture^{27,49,50}, estimated to cause 5–10 million hectares of forest loss annually^{18,51}. Slowing (and, ultimately, ceasing) the expansion of agriculture, particularly into tropical forests, will be an important first step in shifting agriculture onto a more sustainable path.

But will ending the expansion of agriculture negatively affect food supplies? Our analysis suggests that the food production benefits of tropical deforestation are often limited, especially compared to the

environmental damages accrued. First of all, many regions cleared for agriculture in the tropics have low yields compared with their temperate counterparts. The authors of ref. 21 considered crop production and carbon emissions resulting from deforestation and demonstrated that the balance of production gains to carbon losses was often poor in tropical landscapes (Supplementary Fig. 3). Regions of tropical agriculture that do have high yields—particularly areas of sugarcane, oil palm and soybeans—typically do not contribute much to the world's total calorie or protein supplies, especially when crops are used for feed or biofuels. Nevertheless, such crops do provide income, and thereby contribute to poverty alleviation and food security to some sectors of the population.

Although ceasing the expansion of agriculture into tropical forests might have a negative—but probably small—impact on global crop production, losses can be offset elsewhere in the food system. Agricultural production potential that is 'lost' by halting deforestation could be offset by reducing losses of productive farmland and improving yields on existing croplands. Though the 'indirect land use' effects of biofuel production are thought to increase pressure on tropical forests⁵², it may also be true that increasing food production in non-tropical zones might reduce pressures on tropical forests.

Economic drivers hold great sway over deforestation^{53–55}. Ecologically friendly economic incentives could play an important part in slowing forest loss: the proposed Reducing Emissions from Deforestation and Degradation (REDD) programme, market certification, and ecotourism all provide opportunities to benefit economically from forest protection⁵⁶.

Close yield gaps

Increasing food production without agricultural expansion implies that we must increase production on our existing agricultural lands. The best places to improve crop yields may be on underperforming landscapes, where yields are currently below average.

Recent analyses^{57,58} have found large yield variations across the world, even among regions with similar growing conditions, suggesting the existence of 'yield gaps' (Supplementary Fig. 4a). Here we define a yield gap as the difference between crop yields observed at any given location and the crop's potential yield at the same location given current agricultural practices and technologies.

Much of the world experiences yield gaps (Supplementary Fig. 4a) where productivity may be limited by management. There are significant opportunities to increase yields across many parts of Africa, Latin America and Eastern Europe, where nutrient and water limitations seem to be strongest (Supplementary Fig. 4b). Better deployment of existing crop varieties with improved management should be able to close many yield gaps⁵⁹, while continued improvements in crop genetics will probably increase potential yields into the future.

Closing yield gaps could substantially increase global food supplies. Our analysis shows that bringing yields to within 95% of their potential for 16 important food and feed crops could add 2.3 billion tonnes (5×10^{15} kilocalories) of new production, a 58% increase (Fig. 3). Even if yields for these 16 crops were brought up to only 75% of their potential, global production would increase by 1.1 billion tonnes (2.8×10^{15} kilocalories), a 28% increase. Additional gains in productivity, focused on increasing the maximum yield of key crops, are likely to be driven by genetic improvements^{60,61}. Significant opportunities may also exist to improve yield and the resilience of cropping systems by improving 'orphan crops' (such crops have not been genetically improved or had much investment) and preserving crop diversity, which have received relatively little investment to date.

To close global yield gaps, the interwoven challenges of production and environment must again be addressed: conventional approaches to intensive agriculture, especially the unbridled use of irrigation and fertilizers, have been major causes of environmental degradation. Closing yield gaps without environmental degradation will require new approaches, including reforming conventional agriculture and adopting lessons from organic systems and precision agriculture. In addition, closing yield gaps will

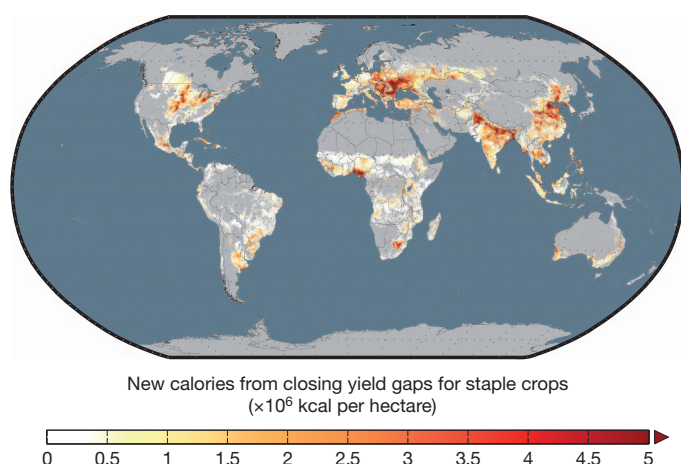


Figure 3 | Closing global yield gaps. Many agricultural lands do not attain their full yield potential. The figure shows the new calories that would be made available to the world from closing the yield gaps for 16 major crops: barley, cassava, groundnut, maize, millet, potato, oil palm, rapeseed, rice, rye, sorghum, soybean, sugarbeet, sugarcane, sunflower and wheat. This analysis shows that bringing the world's yields to within 95% of their potential for these 16 important food and feed crops could add 2.3 billion tonnes (5×10^{13} kilocalories) of new crop production, representing a 58% increase. These improvements in yield can be largely accomplished by improving the nutrient and water supplies to crops in low-yielding regions; further enhancement of global food production could be achieved through improved crop genetics. The methods used to calculate yield gaps and limiting factors are described in the Supplementary Information.

require overcoming considerable economic and social challenges, including the distribution of agricultural inputs and seed varieties and improving market infrastructure.

Increase agricultural resource efficiency

Moving forward, we must find more sustainable pathways for intensification that increase crop production while greatly reducing unsustainable uses of water, nutrients and agricultural chemicals.

Irrigation is currently responsible for water withdrawals of about 2,800 km³ per year from groundwater, lakes and rivers. Irrigation is used on about 24% of croplands and is responsible for delivering 34% of agricultural production¹⁷. In fact, without irrigation, global cereal production would decrease by an estimated 20% (ref. 17), so more land would be required to produce the same amount of food.

However, the benefits and impacts of irrigation are not evenly distributed. Water needed for crop production varies greatly across the world (Supplementary Fig. 5). We find that, when irrigated, 16 staple crops use an average of 0.3 litres per kilocalorie (not including water losses). However, these water requirements are skewed: 80% of irrigated crops require less than 0.4 litres per kilocalorie, while the remaining 20% require 0.7 litres per kilocalorie or more.

Where water is scarce, good water and land management practices can increase irrigation efficiency. For example, curtailing off-field evaporative losses from water storage and transport and reducing on-field losses through mulching and reduced tillage will increase the value of irrigation water.

Chemical fertilizers, manure and leguminous crops have also been key to agricultural intensification. However, they have also led to widespread nutrient pollution and the degradation of lakes, rivers and coastal oceans. In addition, the release of nitrous oxide from fertilized fields contributes to climate change. Excess nutrients also incur energy costs associated with converting atmospheric nitrogen and mining phosphorus^{22,62}.

Even though excess nutrients cause environmental problems in some parts of the world, insufficient nutrients are a major agronomic problem in others. Many yield gaps are mainly due to insufficient nutrient availability (Supplementary Fig. 4b). This 'Goldilocks' problem of nutrients

(that is, there are many regions with too much or too little fertilizer but few that are 'just right') is one of the key issues facing agriculture today⁶³.

Building on recent analyses of crop production, fertilizer use and nutrient cycling^{15,22,64,65}, we examine patterns of agricultural nitrogen and phosphorus balance across the world. Specifically, we show areas of excess nutrients resulting from imbalances between nutrient inputs (fertilizers, legumes and atmospheric deposition), harvest removal and environmental losses (Supplementary Fig. 6). We further analyse the efficiency of nutrient use by comparing applied nutrients to yield for 16 major crops (Supplementary Fig. 6c, d).

Our analysis reveals 'hotspots' of low nutrient use efficiency (Supplementary Fig. 6c, d) and large volumes of excess nutrients (Supplementary Fig. 6e, f). Nutrient excesses are especially large in China⁶⁶, Northern India, the USA and Western Europe. We also find that only 10% of the world's croplands account for 32% of the global nitrogen surplus and 40% of the phosphorus surplus. Targeted policy and management in these regions could improve the balance between yields and the environment. Such actions include reducing excessive fertilizer use, improving manure management, and capturing excess nutrients through recycling, wetland restoration and other practices.

Taken together, these results illustrate many opportunities to improve the water and nutrient efficiency of agriculture without reducing food production. Targeting particular 'hotspots' of low efficiency, measured as the disproportionate use of water and nutrient inputs relative to production, could significantly reduce the environmental problems of intensive agriculture. Furthermore, agroecological innovations in crop and soil management^{1,67} show great promise for improving the resource efficiency of agriculture, maintaining the benefits of intensive agriculture while greatly reducing harm to the environment.

Increase food delivery by shifting diets and reducing waste

While improving crop yields and reducing agriculture's environmental impacts will be instrumental in meeting future needs, it is also important to remember that more food can be delivered by changing our agricultural and dietary preferences. Simply put, we can increase food availability (in terms of calories, protein and critical nutrients) by shifting crop production away from livestock feed, bioenergy crops and other non-food applications.

In Supplementary Fig. 7, we compare intrinsic food production (calories available if all crops were consumed by humans) and delivered food production (calories available based on today's allocation of crops to food, animal feed, and other products, assuming standard conversion factors) for 16 staple crops. By subtracting these two figures, we estimate the potential to increase food supplies by closing the 'diet gap': shifting 16 major crops to 100% human food could add over a billion tonnes to global food production (a 28% increase), or the equivalent of 3×10^{15} food kilocalories (a 49% increase) (Fig. 4).

Of course, the current allocation of crops has many economic and social benefits, and this mixed use is not likely to change completely. But even small changes in diet (for example, shifting grain-fed beef consumption to poultry, pork or pasture-fed beef) and bioenergy policy (for example, not using food crops as biofuel feedstocks) could enhance food availability and reduce the environmental impacts of agriculture.

A large volume of food is never consumed but is instead discarded, degraded or consumed by pests along the supply chain. A recent FAO study⁶⁸ suggests that about one-third of food is never consumed; others⁶⁹ have suggested that as much as half of all food grown is lost; and some perishable commodities have post-harvest losses of up to 100% (ref. 70). Developing countries lose more than 40% of food post-harvest or during processing because of storage and transport conditions. Industrialized countries have lower producer losses, but at the retail or consumer level more than 40% of food may be wasted⁶⁸.

In short, reducing food waste and rethinking dietary, bioenergy and other agricultural choices could substantially improve the delivery of calories and nutrition with no accompanying environmental harm. While wholesale conversions of the human diet and the elimination of

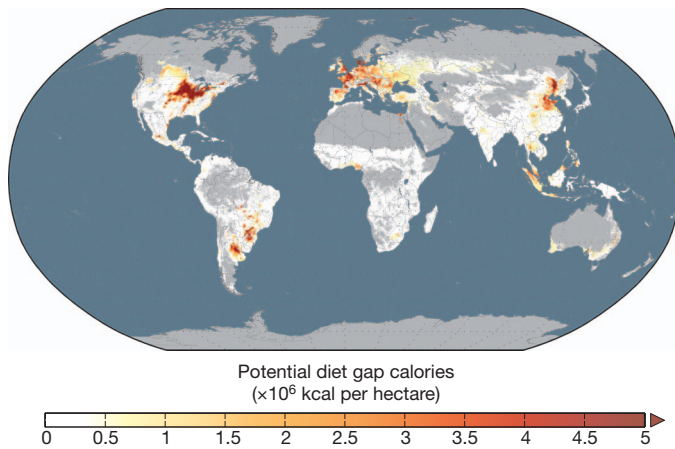


Figure 4 | Closing the diet gap. We estimate the potential to increase food supplies by closing the ‘diet gap’: shifting 16 major crops to 100% human food and away from the current mix of uses (see Fig. 1) could add over a billion tonnes to global food production (a 28% increase for those 16 crops), the equivalent of $\sim 3 \times 10^{15}$ kilocalories more food to the global diet (a 49% increase in food calories delivered).

food waste are not realistic goals, even incremental steps could be extremely beneficial. Furthermore, targeted efforts—such as reducing waste in our most resource-intensive foods, especially meat and dairy—could be designed for optimal impact.

Searching for practical solutions

Today, humans are farming more of the planet than ever, with higher resource intensity and staggering environmental impacts, while diverting an increasing fraction of crops to animals, biofuels and other non-food uses. Meanwhile, almost a billion people are chronically hungry. This must not continue: the requirements of current and future generations demand that we transform agriculture to meet the twin challenges of food security and environmental sustainability.

Our analysis demonstrates that four core strategies can—in principle—meet future food production needs and environmental challenges if deployed simultaneously. Adding them together, they increase global food availability by 100–180%, meeting projected demands while lowering greenhouse gas emissions, biodiversity losses, water use and water pollution. However, all four strategies are needed to meet our global food production and environmental goals; no single strategy is sufficient.

We have described general approaches to solving global agricultural challenges, but much work remains to translate them into action. Specific land use, agricultural and food system tactics must be developed and deployed. Fortunately, many such tactics already exist, including precision agriculture, drip irrigation, organic soil remedies, buffer strips and wetland restoration, new crop varieties that reduce needs for water and fertilizer, perennial grains and tree-cropping systems, and paying farmers for environmental services. However, deploying these tactics effectively around the world requires numerous economic and governance challenges to be overcome. For example, reforming global trade policies, including eliminating price-distorting subsidies and tariffs, will be vital to achieving our strategies.

In developing improved land use and agricultural practices, we recommend following these guidelines:

- (1) Solutions should focus on critical biophysical and economic ‘leverage points’ in agricultural systems, where major improvements in food production or environmental performance may be achieved with the least effort and cost.
- (2) New practices must also increase the resilience of the food system. High-efficiency, industrialized agriculture has many benefits, but it is vulnerable to disasters⁷¹, including climatic disturbances, new diseases and economic calamities.

(3) Agricultural activities have many costs and benefits, but methods of evaluating the trade-offs are still poorly developed⁷². We need better data and decision support tools to improve management decisions⁷³, productivity and environmental stewardship.

(4) The search for agricultural solutions should remain technology-neutral. There are multiple paths to improving the production, food security and environmental performance of agriculture, and we should not be locked into a single approach a priori, whether it be conventional agriculture, genetic modification or organic farming.

The challenges facing agriculture today are unlike anything we have experienced before, and they require revolutionary approaches to solving food production and sustainability problems. In short, new agricultural systems must deliver more human value, to those who need it most, with the least environmental harm.

Published online 12 October 2011.

1. International Assessment of Agricultural Knowledge (IAASTD). *Agriculture at a Crossroads, Global Report* Chs 1, 4 (Island Press, 2009); [http://www.agassessment.org/reports/IAASTD/EN/Agriculture at a Crossroads_Global Report \(English\).pdf](http://www.agassessment.org/reports/IAASTD/EN/Agriculture%20at%20a%20Crossroads_Global%20Report%20(English).pdf).
2. *Reaping the Benefits: Science and the Sustainable Intensification of Global Agriculture* 1–10, 47–50 (The Royal Society, 2009); <http://royalsociety.org/Reapingthebenefits/>.
3. Pelletier, N. & Tyedmers, P. Forecasting potential global environmental costs of livestock production 2000–2050. *Proc. Natl Acad. Sci. USA* **107**, 18371–18374 (2010).
4. Food and Agriculture Organization of the United Nations (FAO). *The State of Food Insecurity in the World: Economic crises—Impacts and Lessons Learned* 8–12 (FAO, 2009).
5. Thurow, R. & Kilman, S. *Enough: Why the World’s Poorest Starve in an Age of Plenty* Chs 2, 4, 12 (Perseus Books, 2009).
6. Godfray, H. C. J. *et al.* Food security: the challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010).
- This article reviews a recent effort by the UK-based Foresight Project, which assessed global conditions and trends in agriculture and food security, and set the benchmark for the world’s discussions on this important topic.**
7. Naylor, R. Expanding the boundaries of agricultural development. *Food Security* **3**, 233–251 (2011).
8. Kearney, J. Food consumption trends and drivers. *Phil. Trans. R. Soc. B* **365**, 2793–2807 (2010).
9. Cirera, X. & Masset, E. Income distribution trends and future food demand. *Phil. Trans. R. Soc. B* **365**, 2821–2834 (2010).
10. Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
- This paper reviews the global extent of land use practices, especially agriculture, and how it has become a transformative force in the global environment—through changes in climate, water resources, biogeochemical cycles and biodiversity.**
11. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being Vol. 2 Scenarios: Findings of the Scenarios Working Group* Ch. 9 (Island Press, 2005).
12. Power, A. G. Ecosystem services and agriculture: tradeoffs and synergies. *Phil. Trans. R. Soc. B* **365**, 2959–2971 (2010).
13. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
- This article presents a new way of thinking about the condition of the global environment and the idea of “planetary boundaries”—points where more environmental deterioration may “tip” the global environment far out of the current condition.**
14. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Glob. Biogeochem. Cycles* **22**, GB1003 (2008).
15. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Glob. Biogeochem. Cycles* **22**, GB1022 (2008).
16. Portmann, F. T., Siebert, S. & Döll, P. MIRCA 2000: global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. *Glob. Biogeochem. Cycles* **24**, GB1011 (2010).
17. Siebert, S. & Döll, P. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *J. Hydrol.* **384**, 198–217 (2010).
- This paper presents a state-of-the-art global assessment of how water resources (both ‘blue’ and ‘green’ water) are deployed in agriculture, primarily through irrigation, and how this is related to food production.**
18. Food and Agriculture Organization of the United Nations (FAOSTAT). <http://faostat.fao.org/site/567/default.aspx#ancor> (accessed, March 2011).
19. Ramankutty, N., Foley, J. A., Norman, J. & McSweeney, K. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Glob. Ecol. Biogeogr.* **11**, 377–392 (2002).
20. Ellis, E. C., Klein Goldewijk, K., Siebert, S., Lightman, D. & Ramankutty, N. Anthropogenic transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.* **19**, 589–606 (2010).

21. West, P. C. *et al.* Trading carbon for food: global comparison of carbon stocks vs. crop yields on agricultural land. *Proc. Natl Acad. Sci. USA* **107**, 19645–19648 (2010).
This paper explores how future expansion of agriculture would lead to increasing greenhouse gas emissions (from deforestation) and increasing food production (by adding more farmland), and assesses the geographic patterns of the tradeoffs between the two.
22. MacDonald, G. K., Bennett, E. M., Potter, P. A. & Ramankutty, N. Agronomic phosphorus imbalances across the world's croplands. *Proc. Natl Acad. Sci. USA* **108**, 3086–3091 (2011).
23. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671–677 (2002).
24. Steinfeld, H. *et al.* *Livestock's Long Shadow: Environmental Issues and Options* 1–20 (FAO, 2006).
25. Ramankutty, N. & Foley, J. A. Estimating historical changes in global land cover: croplands from 1700 to 1992. *Glob. Biogeochem. Cycles* **13**, 997–1027 (1999).
26. Gibbs, H. *et al.* Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl Acad. Sci. USA* **107**, 16732–16737 (2010).
27. Foley, J. A. *et al.* Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. *Front. Ecol. Environ.* **5**, 25–32 (2007).
28. Friedlingstein, P. *et al.* Update on CO₂ emissions. *Nature Geosci.* **3**, 811–812 (2010).
29. DeFries, R. & Rosenzweig, C. Toward a whole-landscape approach for sustainable land use in the tropics. *Proc. Natl Acad. Sci. USA* **107**, 19627–19632 (2010).
30. Rosegrant, M. W., Cai, X. & Cline, S. A. *World Water and Food to 2025: Dealing with Scarcity* 1–32 (International Food Policy Research Institute, 2002).
31. Gleick, P. H. Global freshwater resources: soft-path solutions for the 21st century. *Science* **302**, 1524–1528 (2003).
32. Matson, P., Parton, W., Power, A. & Swift, M. Agricultural intensification and ecosystem properties. *Science* **277**, 504–509 (1997).
33. Tilman, D. *et al.* Forecasting agriculturally driven global environmental change. *Science* **292**, 281–284 (2001).
34. Vorosmarty, C. J., Green, P., Salisbury, J. & Lammers, R. B. Global water resources: vulnerability from climate change and population growth. *Science* **289**, 284–288 (2000).
35. Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321**, 926–929 (2008).
36. Gleick, P. H., Cooley, H. & Morikawa, M. *The World's Water 2008–2009: The Biennial Report on Freshwater Resources* (eds Gleick, P. H. *et al.*) 202–210 (Island Press, 2009).
37. Postel, S. L., Daily, G. C. & Ehrlich, P. R. Human appropriation of renewable fresh water. *Science* **271**, 785–788 (1996).
38. Gordon, L. J. *et al.* Human modification of global water vapor flows from the land surface. *Proc. Natl Acad. Sci. USA* **102**, 7612–7617 (2005).
39. Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J. M. Human domination of Earth's ecosystems. *Science* **277**, 494–499 (1997).
40. Smil, V. Phosphorus in the environment: natural flows and human interferences. *Annu. Rev. Energy Environ.* **25**, 53–88 (2000).
41. Bennett, E. M., Carpenter, S. R. & Caraco, N. F. Human impact on erodable phosphorus and eutrophication: a global perspective. *Bioscience* **51**, 227–234 (2001).
42. Canfield, D. E., Glazer, A. N. & Falkowski, P. G. The evolution and future of earth's nitrogen cycle. *Science* **330**, 192–196 (2010).
43. Galford, G. L. *et al.* Greenhouse gas emissions from alternative futures of deforestation and agricultural management in the southern Amazon. *Proc. Natl Acad. Sci. USA* **107**, 19649–19654 (2010).
44. van der Werf, G. *et al.* CO₂ emissions from forest loss. *Nature Geosci.* **2**, 737–738 (2009).
45. Canadell, J. G. *et al.* Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl Acad. Sci. USA* **104**, 18866–18870 (2007).
46. Vergé, X., De Kimpe, C. & Desjardins, R. Agricultural production, greenhouse gas emissions and mitigation potential. *Agric. For. Meteorol.* **142**, 255–269 (2007).
47. DeFries, R. S., Foley, J. A. & Asner, G. P. Land-use choices: balancing human needs and ecosystem function. *Front. Ecol. Environ.* **2**, 249–257 (2004).
48. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2007: IPCC Fourth Assessment Report (AR4)* (Cambridge University Press, 2007).
49. Gibbs, H. K. *et al.* Carbon payback times for crop-based biofuel expansion in the tropics: the effects of changing yield and technology. *Environ. Res. Lett.* **3**, 034001 (2008).
50. Fargione, J., Hill, J., Tilman, D., Polasky, S. & Hawthorne, P. Land clearing and the biofuel carbon debt. *Science* **319**, 1235–1238 (2008).
51. Mayaux, P. *et al.* Tropical forest cover change in the 1990s and options for future monitoring. *Phil. Trans. R. Soc. B* **360**, 373–384 (2005).
52. Searchinger, T. *et al.* Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **319**, 1238–1240 (2008).
53. Lambin, E. F. & Meyfroidt, P. Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl Acad. Sci. USA* **108**, 3465–3472 (2011).
54. Rudel, T. K. *et al.* Agricultural intensification and changes in cultivated areas, 1970–2005. *Proc. Natl Acad. Sci. USA* **106**, 20675–20680 (2009).
55. DeFries, R. S., Rudel, T., Uriarte, M. & Hansen, M. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geosci.* **3**, 178–181 (2010).
56. Kremen, C., Daily, G. C., Klein, A. & Scofield, D. Inadequate assessment of the ecosystem service rationale for conservation: reply to Ghazoul. *Conserv. Biol.* **22**, 795–798 (2008).
57. Licker, R. *et al.* Mind the gap: how do climate and agricultural management explain the 'yield gap' of croplands around? *Global Ecol. Biogeogr.* **19**, 769–782 (2010).
These authors present a new technique for estimating global patterns of yield and 'yield gaps', highlighting opportunities for improving food production around the world.
58. Neumann, K., Verburg, P. H., Stehfest, E. & Müller, C. The yield gap of global grain production: a spatial analysis. *Agric. Syst.* **103**, 316–326 (2010).
59. Sánchez, P. A. Tripling crop yields in tropical Africa. *Nature Geosci.* **3**, 299–300 (2010).
60. Jaggard, K. W., Qi, A. & Ober, E. S. Possible changes to arable crop yields by 2050. *Phil. Trans. R. Soc. B* **365**, 2835–2851 (2010).
61. Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **327**, 818–822 (2010).
62. Cordell, D., Drangert, J. O. & White, S. The story of phosphorus: global food security and food for thought. *Glob. Environ. Change* **19**, 292–305 (2009).
63. Cassman, K. G., Dobermann, A. & Walters, D. T. Agroecosystems, nitrogen-use efficiency, and nitrogen management. *Ambio* **31**, 132–140 (2002).
64. Potter, P., Ramankutty, N., Bennett, E. M. & Donner, S. D. Characterizing the spatial patterns of global fertilizer application and manure production. *Earth Interact.* **14**, 1–22 (2010).
65. Liu, J. *et al.* A high-resolution assessment on global nitrogen flows in cropland. *Proc. Natl Acad. Sci. USA* **107**, 8035–8040 (2010).
66. Vitousek, P. *et al.* Nutrient imbalances in agricultural development. *Science* **324**, 1519–1520 (2009).
67. Chen, X. P. *et al.* Integrated soil-crop system management for food security. *Proc. Natl Acad. Sci. USA* **108**, 6399–6404 (2011).
68. Gustavsson, J., Cederberg, C., Sonesson, U., van Otterdijk, R. & Meybeck, A. *Global Food Losses and Food Waste* Section 3.2 (Study conducted for the International Congress "Save Food!" at Interpack2011, Düsseldorf, Germany) (FAO, Rural Infrastructure and Agro-Industries Division, 2011).
69. Lundqvist, J., De Fraiture, C. & Molden, D. *Saving Water: from Field to Fork: Curbing Losses and Waste in the Food Chain* 20–23 (Stockholm International Water Institute, 2008).
70. Parfitt, J., Barthel, M. & Macnaughton, S. Food waste within food supply chains: quantification and potential for change to 2050. *Phil. Trans. R. Soc. B* **365**, 3065–3081 (2010).
71. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proc. Natl Acad. Sci. USA* **106**, 15594–15598 (2009).
72. Sachs, J. *et al.* Monitoring the world's agriculture. *Nature* **466**, 558–560 (2010).
73. Zaks, D. P. M. & Kucharik, C. J. Data and monitoring needs for a more ecological agriculture. *Environ. Res. Lett.* **6**, 014017 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for the support of NASA and the National Science Foundation. We also acknowledge the support of the Stockholm Resilience Centre, for convening a workshop on meeting global agricultural demands while staying within the 'planetary limits'. We thank C. Godfray and C. Prentice for comments on the manuscript. We also thank M. Hoff and S. Karnas for help with the manuscript and figures.

Author Contributions J.A.F., N.R., K.A.B., E.S.C., J.S.G., M.J., N.D.M., C.O'C., D.K.R. and P.C.W. conducted most of the data production, analysis and shared writing responsibilities. C.B., C.M., S.S. and D.T. contributed data and shared in the scoping and writing responsibilities. E.M.B., S.R.C., J.H., S.P., J.R., J.S. and D.P.M.Z. shared in the scoping and writing responsibilities.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.F. (jfoley@umn.edu).

Structural basis for the bifunctionality of fructose-1,6-bisphosphate aldolase/phosphatase

Shinya Fushinobu^{1*}, Hiroshi Nishimasu^{2*}, Daiki Hattori¹, Hyun-Jin Song¹ & Takayoshi Wakagi¹

Enzymes catalyse specific reactions and are essential for maintaining life. Although some are referred to as being bifunctional, they consist of either two distinct catalytic domains or a single domain that displays promiscuous substrate specificity¹. Thus, one enzyme active site is generally responsible for one biochemical reaction. In contrast to this conventional concept, archaeal fructose-1,6-bisphosphate (FBP) aldolase/phosphatase (FBPA/P) consists of a single catalytic domain, but catalyses two chemically distinct reactions of gluconeogenesis: (1) the reversible aldol condensation of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GA3P) to FBP; (2) the dephosphorylation of FBP to fructose-6-phosphate (F6P)². Thus, FBPA/P is fundamentally different from ordinary enzymes whose active sites are responsible for a specific reaction. However, the molecular mechanism by which FBPA/P achieves its unusual bifunctionality remains unknown. Here we report the crystal structure of FBPA/P at 1.5-Å resolution in the aldolase form, where a critical lysine residue forms a Schiff base with DHAP. A structural comparison of the aldolase form with a previously determined phosphatase form³ revealed a dramatic conformational change in the active site, demonstrating that FBPA/P metamorphoses its active-site architecture to exhibit dual activities. Thus, our findings expand the conventional concept that one enzyme catalyses one biochemical reaction.

FBPA/P was initially identified as class V fructose-1,6-bisphosphatase (FBPase)⁴, and is responsible for gluconeogenesis in the hyperthermophilic archaeon *Thermococcus kodakaraensis*⁵. Class V FBPases lack sequence homology with other FBPase or FBP aldolase (FBPA) proteins, and are found in the genomes of virtually all Archaea and deeply branching Bacteria (Supplementary Fig. 1). FBPA/P shows higher activity in the condensation reaction than in the FBP cleavage reaction² (Supplementary Fig. 2), indicating its involvement in gluconeogenesis, rather than in glycolysis. Thus, FBPA/P has been proposed to represent an ancestral gluconeogenic enzyme that ensures unidirectional gluconeogenesis in chemolithoautotrophic organisms². We previously determined the crystal structure of an FBPA/P protein from *Sulfolobus tokodaii* (ST0318) at 1.8-Å resolution, in complex with FBP and catalytically essential Mg²⁺ ions, which represents the 'FBPase form'³. The structure revealed a unique fold and lacked similarity to those of other known FBPase and FBPA proteins. Unexpectedly, biochemical experiments indicated that a conserved lysine residue (Lys 232 of ST0318), which is located away from the bound FBP molecule in the ST0318 structure, forms a Schiff base with DHAP during the FBPA reaction², suggesting that FBPA/P undergoes a large structural change for its dual activities. However, the molecular mechanism by which FBPA/P exhibits its bifunctionality remains unknown. Here, we present the crystal structure of ST0318 in the DHAP-Schiff base intermediate state, which represents the 'FBPA form'. A comparison of the high-resolution structures of FBPA/P in the aldolase and phosphatase forms provides the structural basis for its bifunctionality.

We measured the FBPA and FBPase activities of the purified recombinant ST0318 enzyme, and confirmed that the enzyme exhibits both activities (Table 1), demonstrating that ST0318 is also an FBPA/P enzyme. We crystallized ST0318 in the presence of DHAP and Mg²⁺, and determined the crystal structure at 1.5-Å resolution (Fig. 1a, Supplementary Fig. 3a and Supplementary Table 1). The crystal belongs to the space group *I*422, and is isomorphous to the previously determined FBPase form of ST0318 (PDB code 1UMG). The monomer structures are almost identical in the FBPA and FBPase forms, with root mean square deviations of 3.28 and 0.29 Å for the Cα atoms of the overall polypeptide and the regions without the three mobile loops (residues 97–110, 219–233 and 346–361). As in the FBPase form³, the protomers in the asymmetric units are related by crystallographic symmetry to form an octamer (Supplementary Fig. 4). Thus, the ternary and quaternary structures are virtually the same in these two forms. We observed an electron density for the bound DHAP in the active site, located at the dimer interface of the ring-shaped tetramer (Supplementary Fig. 5). As previously postulated², Lys 232 formed a Schiff base intermediate with DHAP (Fig. 1a and Supplementary Fig. 3a). The Lys 232 Nζ atom and the DHAP C1, C2 and C3 atoms adopt a nearly planar arrangement, indicating that the intermediate is similar to the imine (iminium) form observed in a typical class I FBPA, for example rabbit muscle FBPA (rmFBPA)⁶. The intermediate may have been trapped, owing to the absence of the second substrate, GA3P. The DHAP phosphate group coordinates three Mg²⁺ ions (Mg2–Mg4) (Supplementary Fig. 6), whereas the FBP 1-phosphate group coordinates four Mg²⁺ ions (Mg1–Mg4) in the FBPase form³. The DHAP hydroxyl group is recognized by Arg266 and Asp287 through direct hydrogen bonds, and by Gln242' (the prime symbol indicates residues from the neighbouring protomer) through a water-mediated hydrogen bond. These hydrogen-bonding interactions are likely to be critical for the discrimination between DHAP and GA3P, as GA3P has a carbonyl group and lacks a hydroxyl

Table 1 | Activities of wild type and mutants of ST0318

Enzyme	k_{cat} (s ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (s ⁻¹ mM ⁻¹)
FBPase			
Wild type	0.62 ± 0.02	0.027 ± 0.003	23
Y229F	0.66 ± 0.02	0.027 ± 0.003	25
Y348F	0.26 ± 0.01	0.036 ± 0.002	7.2
FBPA (anabolic direction)			
Wild type	0.91 ± 0.04	0.19 ± 0.02	4.7
Y229F	ND*	–	–
Y348F	0.10 ± 0.01	0.34 ± 0.04	0.29
FBPA (catabolic direction)			
Wild type	0.027 ± 0.011†	–	–
Y229F	ND*	–	–
Y348F	0.026 ± 0.004†	–	–

* ND, not detected.

† Enzyme activity was measured with 5 mM FBP.

¹Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. ²Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.

*These authors contributed equally to this work.

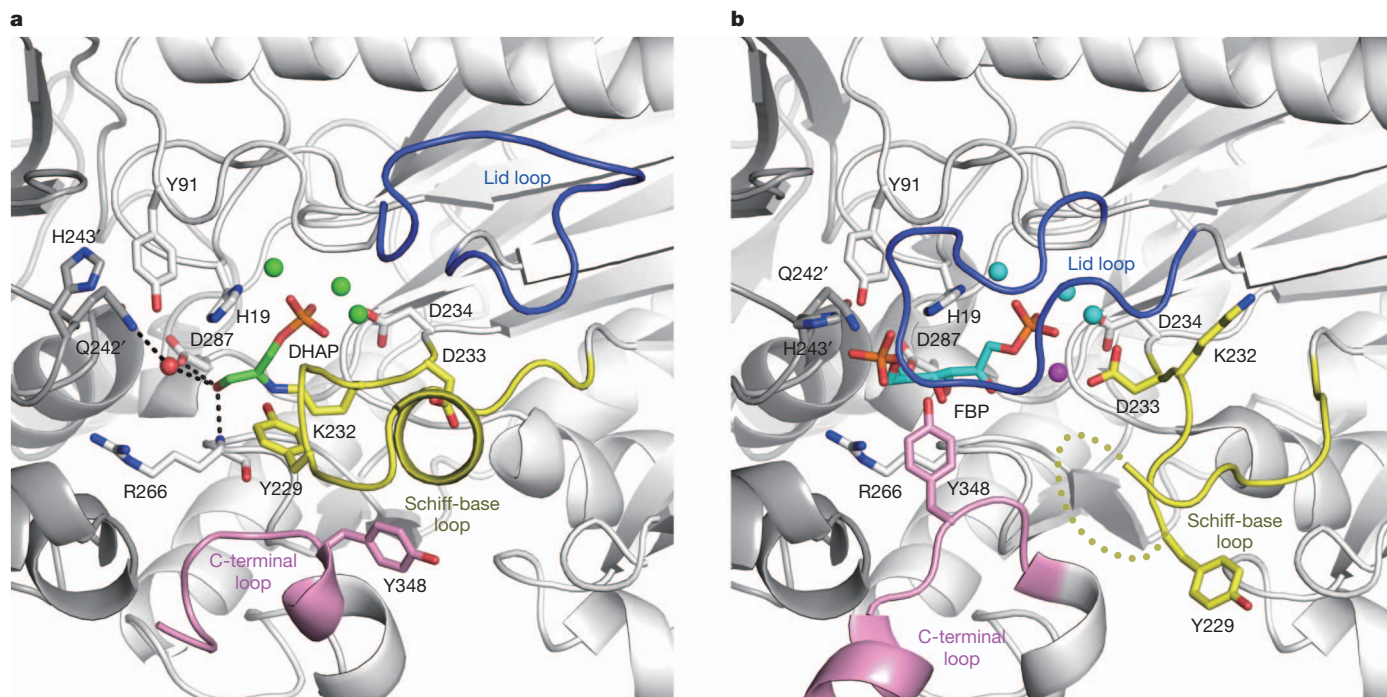


Figure 1 | Active sites of ST0318 in the two forms. **a, b**, DHAP–Schiff base complex (the FBPA form) (**a**) and FBP complex (the FBPase form, PDB code 1UMG)³ (**b**). DHAP and FBP are coloured green and cyan, respectively. The neighbouring protomer is coloured dark grey, and its residues are labelled with prime symbols. The lid, Schiff-base and C-terminal loops are coloured blue,

yellow and pink, respectively. In (**a**), Mg^{2+} ions (Mg2–4) and a water molecule are shown as green and red spheres, respectively. Hydrogen bonds are shown as black dashed lines. In (**b**), Mg^{2+} ions (Mg1 and Mg2–4) are shown as magenta and cyan spheres, respectively. The disordered region is indicated by a yellow dashed line.

group at the C3 position. The enzyme displayed the activity under the crystallization conditions (albeit approximately 50% compared with that under standard conditions), and the dissolved crystals exhibited activity comparable to that of the purified enzyme, confirming that the FBPA/P–DHAP complex structure determined here represents a genuine reaction intermediate. The DHAP molecule is located at a position similar to that of the FBP molecule in the FBPase form, demonstrating that FBPA/P catalyses both the FBPA and FBPase reactions at a single site.

A structural comparison of the FBPA and FBPase forms reveals striking differences in three loop regions at the active site: a lid loop (residues 97–110), a Schiff-base loop (residues 219–233) that contains the Schiff base-forming Lys232 residue, and a carboxy (C)-terminal loop (residues 346–361) (Fig. 1). Although we could confidently trace the polypeptide chain, except for part of the C-terminal loop (residues 354–360), the loop regions exhibit relatively higher average B-factors than that for the overall polypeptide (Supplementary Table 2), indicating their flexible nature. In the FBPase form, the lid and C-terminal loops interact with FBP, whereas the Schiff-base loop is further away from the active site and does not contact FBP³ (Fig. 1b and Supplementary Fig. 3b). In the FBPA form, the lid and C-terminal loops are displaced outwards from the active site and do not contact DHAP, whereas the Schiff-base loop cuts into the active site, allowing Tyr229 and Lys232 to interact with DHAP (Fig. 1a and Supplementary Fig. 3a). Although Asp234 similarly holds Mg2 in the two forms, Asp233, which holds the catalytically essential Mg1 in the FBPase form³, is flipped out in the FBPA form, resulting in the dissociation of Mg1.

A structural comparison of ST0318 in the two forms with rmFBPA in the DHAP- and FBP–Schiff base intermediate states^{6,7} reveals an unexpected similarity in the active-site configurations between FBPA/P and class I FBPA (Fig. 2a and Supplementary Fig. 7a). Notably, the covalent DHAP intermediates of ST0318 and rmFBPA superimpose well, even though the C α atoms of their Schiff base-forming lysine residues are far apart, owing to their distinct overall structures

(Supplementary Fig. 8). Based on this analogy, we propose a mechanism for FBPA/P-catalysed reactions (Fig. 3). The aldol condensation reaction consists of dehydration, carbanion formation, C3–C4 bond formation and hydration steps, in which general acid/base catalysts participate. In rmFBPA, Glu 187 reportedly functions as the general acid/base residue in multiple steps⁷. In addition, Asp33 (ref. 8) or Tyr363 (ref. 6) may be involved in the second carbanion formation step, by accepting a proton from the C3 methylene of the DHAP–Schiff base intermediate. Tyr229 and Asp287 in ST0318 are located at positions similar to those of Glu187 and Asp33 in rmFBPA (Fig. 2a and Supplementary Fig. 7a). Tyr229 is located in the vicinity of the Lys232 N ζ (3.2 Å), the DHAP C2 (3.0 Å) and C3 (3.3 Å), and the superposed FBP O4 (2.6 Å) atoms, whereas Asp287 is further away from the DHAP C2 (4.9 Å) and C3 (3.9 Å) atoms (Fig. 2b and Supplementary Fig. 7b). Moreover, Tyr229 is completely conserved in the FBPA/P proteins, whereas Asp287 is replaced with a cysteine residue in the *Thermus thermophilus* and *Cenarchaeum symbiosum* FBPA/Ps (Supplementary Fig. 1). Indeed, the Y229F mutant retained the FBPase activity, but lost the FBPA activity (Table 1). These observations suggest that Tyr229 serves as the catalytic acid/base residue for all steps (Fig. 3). In contrast to the FBPase form, the active site is accessible to the solvent in the FBPA form, and thus seems suitable for the hydration reaction. A comparison of the high-resolution structures of the two forms provides insight into the GA3P recognition mechanism. In the FBPA form, two water molecules form hydrogen bonds with Tyr229 and Asp287, and they are located at positions equivalent to the FBP 4- and 5-hydroxyl groups in the FBPase form, respectively (Fig. 2b and Supplementary Fig. 7b). In addition, the residues involved in the recognition of the FBP 6-phosphate group (His19, Tyr91, Gln242' and His243') are located at similar positions in the two forms. Thus, it is likely that the GA3P phosphate group is recognized by His19, Tyr91, Gln242' and His243', and the GA3P carbonyl and hydroxyl groups are recognized by Tyr229 and Asp287, respectively. The nucleophilic water molecule for the FBPase reaction is already bound to the FBPA form, and is held by Mg2 and Mg3

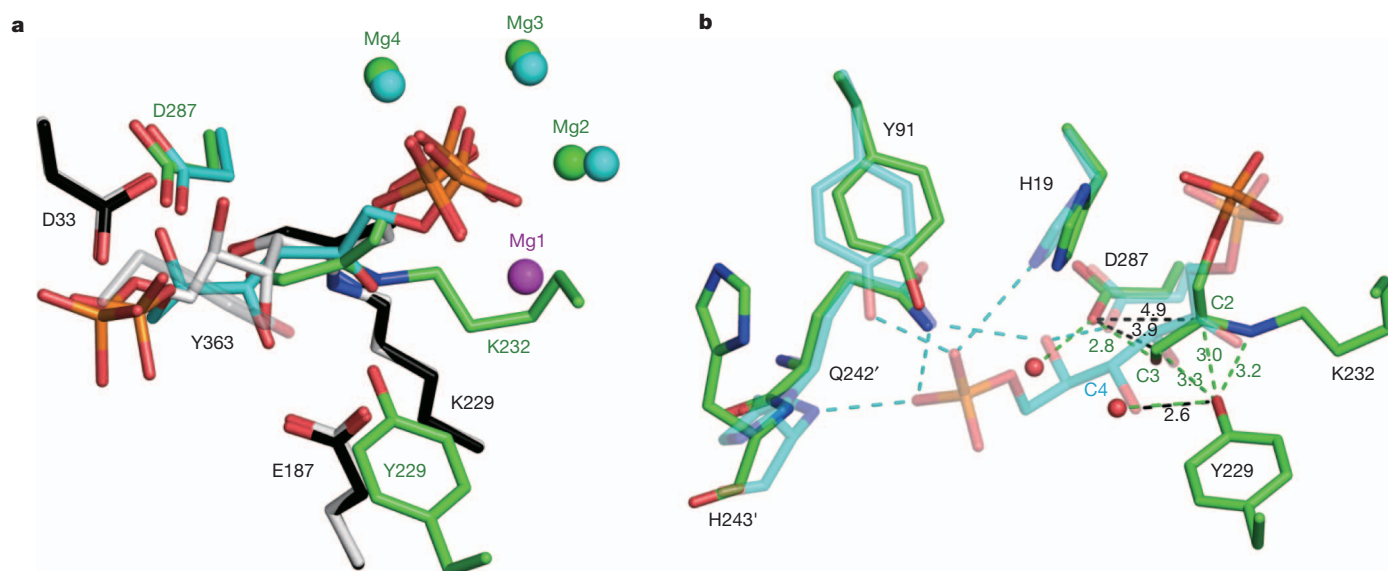


Figure 2 | Catalytic components of FBPA. **a**, Superimposition of the FBPA (green) and FBPase forms (cyan, PDB code 1UMG) of ST0318, and the DHAP-Schiff base (enamine) intermediate (black, PDB code 2QUT) and the FBP-Schiff base intermediate (pericyclic transition state) of wild-type rmFBPA (white, PDB code 1ZAI), and Tyr 363 of the rmFBPA K146M mutant (semi-transparent grey, PDB code 2QUU). The residues of ST0318 and rmFBPA are labelled with green and black characters, respectively. The Mg^{2+} ion (Mg1) in the FBPase form is shown as a magenta sphere. **b**, Superimposition of the FBPA

(green) and FBPase forms (semi-transparent cyan, PDB code 1UMG) of ST0318. Water molecules in the FBPA form are shown as red spheres. The interactions in the FBPA and FBPase forms are depicted by green and cyan dashed lines, respectively, and the distances (in ångströms) between the protein atoms (Tyr 229 or Asp 287) and the DHAP-Schiff base are shown (green). The distances (in ångströms) between Asp 287 and the DHAP C2 and C3 atoms and between Tyr 229 and the FBP O4 atom are also shown with black dashed lines.

(Supplementary Fig. 6). In addition, the catalytic base for the FBPase reaction, Asp 12, is located at a similar position in both forms. However, DHAP is protected from dephosphorylation, probably because of the lack of Mg1, which is required for stabilizing the developing negative charge after phosphate cleavage in the FBPase reaction³. Hydrolysis of the Schiff-base releases the loop containing Lys 232 from the active site, thereby enabling the enzyme to bind Mg1. In addition, the closure of the lid and C-terminal loops would be important for the FBPase reaction, as they stabilize FBP binding. Gly 104 and Asn 105 on the lid loop interact with the FBP 6- and 1-phosphate groups, respectively, whereas Tyr 348 on the C-terminal loop interacts with the FBP 4-hydroxyl and 6-phosphate groups³.

In summary, the crystal structures of FBPA/P in the two forms revealed that FBPA/P achieves its bifunctionality by transforming its active-site architecture, through the toggle switch-like motions of the three mobile loop regions. Conformational fluctuations, for example loop movements and domain motions, are essential for substrate binding and product release in enzyme functions^{9–11}. A recent study showed that a His/Trp biosynthesis isomerase, PriA, exhibits bisubstrate specificity through a substrate-induced metamorphosis of the active-site architecture¹². However, the role of conformational fluctuations in enzyme catalysis and bifunctionality has been less clear. To our knowledge, this study is the first to elucidate the molecular mechanism by which an enzyme catalyses multiple chemical reactions at a single

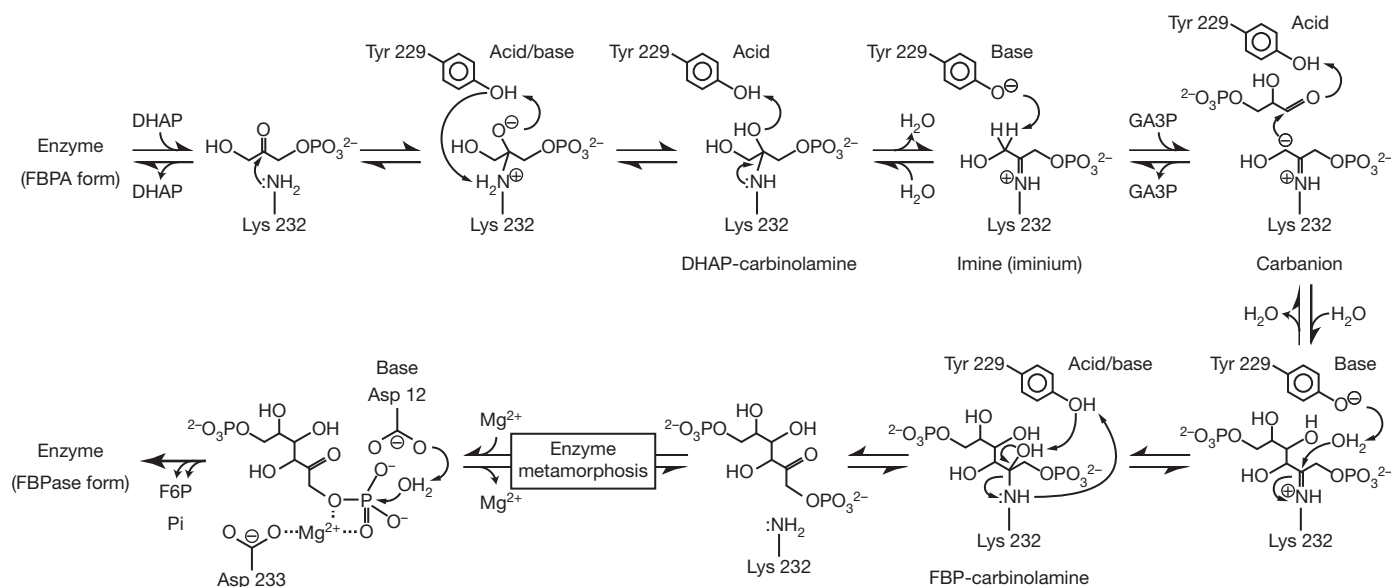


Figure 3 | Proposed mechanism of the FBPA and FBPase reactions catalysed by ST0318. The crystal structure determined here represents the imine (iminium) intermediate.

site. Furthermore, our findings suggest the possible existence of undiscovered enzymes that also catalyse multiple chemical reactions at a single site.

METHODS SUMMARY

The wild-type and mutant ST0318 proteins were expressed in *Escherichia coli* and purified to homogeneity, as described previously³. Crystallization was performed at 25 °C, using the sitting-drop vapour-diffusion method. Crystals were obtained by mixing 1 µl of protein solution, consisting of 14 mg ml⁻¹ ST0318, 20 mM Tris-HCl (pH 7.5), 5 mM DHAP and 5 mM MgCl₂, and 1 µl of reservoir solution, consisting of 0.1 M Bicine-KOH (pH 9.0), 10% PEG 20,000 and 2% dioxane. X-ray diffraction data were collected at the NW12A station ($\lambda = 1.0 \text{ \AA}$) at the Photon Factory AR, High Energy Accelerator Research Organization, Tsukuba, Japan. The crystal structure of ST0318 in the FBPAse form (PDB code 1UMG) was used as the initial model for refinement. Data collection and refinement statistics are provided in Supplementary Table 1. Site-directed mutagenesis was performed with a PrimeSTAR mutagenesis kit (Takara Bio). The FBPA and FBPAse activities were both measured at 48 °C using a coupled spectrophotometric assay, as described previously².

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 15 August 2011.

Published online 9 October 2011.

1. Moore, B. Bifunctional and moonlighting enzymes: lighting the way to regulatory control. *Trends Plant Sci.* **9**, 221–228 (2004).
2. Say, R. F. & Fuchs, G. Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464**, 1077–1081 (2010).
3. Nishimasu, H., Fushinobu, S., Shoun, H. & Wakagi, T. The first crystal structure of the novel class of fructose-1,6-bisphosphatase present in thermophilic archaea. *Structure* **12**, 949–959 (2004).
4. Rashid, N. *et al.* A novel candidate for the true fructose-1,6-bisphosphatase in archaea. *J. Biol. Chem.* **277**, 30649–30655 (2002).

5. Sato, T. *et al.* Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.* **186**, 5799–5807 (2004).
6. St-Jean, M. & Sygusch, J. Stereospecific proton transfer by a mobile catalyst in mammalian fructose-1,6-bisphosphate aldolase. *J. Biol. Chem.* **282**, 31028–31037 (2007).
7. St-Jean, M., Lafrance-Vanasse, J., Liotard, B. & Sygusch, J. High resolution reaction intermediates of rabbit muscle fructose-1,6-bisphosphate aldolase: substrate cleavage and induced fit. *J. Biol. Chem.* **280**, 27262–27270 (2005).
8. Choi, K. H., Shi, J., Hopkins, C. E., Tolan, D. R. & Allen, K. N. Snapshots of catalysis: the structure of fructose-1,6-(bis)phosphate aldolase covalently bound to the substrate dihydroxyacetone phosphate. *Biochemistry* **40**, 13868–13875 (2001).
9. Huber, R. & Bennett, W. S. Jr. Functional significance of flexibility in proteins. *Biopolymers* **22**, 261–279 (1983).
10. Bustamante, C., Chemla, Y. R., Forde, N. R. & Izhaky, D. Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **73**, 705–748 (2004).
11. Tousignant, A. & Pelletier, J. N. Protein motions promote catalysis. *Chem. Biol.* **11**, 1037–1042 (2004).
12. Due, A. V., Kuper, J., Geerlof, A., Kries, J. P. & Wilmanns, M. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc. Natl Acad. Sci. USA* **108**, 3554–3559 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank O. Nureki for discussions, and the staff of the Photon Factory for the X-ray data collection.

Author Contributions S.F. collected the diffraction data and determined the crystal structure. H.N. prepared the wild-type and mutant proteins and crystallized the wild-type protein. D.H. prepared and crystallized the wild-type protein. D.H. and H.-J.S. measured the enzyme activity. S.F., H.N. and T.W. conceived the project and wrote the manuscript. All authors commented on the manuscript.

Author Information The atomic coordinates and structure factors are deposited in Protein Data Bank under accession number 3R1M. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.W. (atwakag@mail.ecc.u-tokyo.ac.jp).

METHODS

Protein preparation and crystallography. Wild-type and mutant ST0318 proteins were expressed in *E. coli* and purified to homogeneity, as described previously³. Crystallization was performed at 25 °C, using the sitting-drop vapour-diffusion method. Crystals were obtained by mixing 1 µl of protein solution, consisting of 14 mg ml⁻¹ ST0318, 20 mM Tris-HCl (pH 7.5), 5 mM DHAP and 5 mM MgCl₂, and 1 µl of reservoir solution, consisting of 0.1 M Bicine-KOH (pH 9.0), 10% PEG 20,000 and 2% dioxane. X-ray diffraction data were collected at the NW12A station ($\lambda = 1.0 \text{ \AA}$) at the Photon Factory AR, High Energy Accelerator Research Organization, Tsukuba, Japan. Crystals were cryoprotected in the reservoir solution supplemented with 25% 2-methyl-2,4-pentanediol, and were flash-cooled at 100 K in a stream of nitrogen gas. Data were processed using HKL2000 (ref. 13). The previously determined ST0318 structure (PDB code 1UMG) was used as the initial model for refinement. Manual model rebuilding and refinement were performed using Coot¹⁴ and Refmac5 (ref. 15). The final model contains residues 2–353 and 361–364, 264 water molecules, one 2-methyl-2,4-pentanediol, three Mg²⁺ ions and one DHAP molecule (O2 atom dehydrated) covalently attached to Lys 232. Data collection and refinement statistics are provided in Supplementary Table 1. Molecular graphic images were prepared using PyMol (Delano Scientific).

Site-directed mutagenesis and enzyme assay. Site-directed mutagenesis was performed with a PrimeSTAR mutagenesis kit (Takara Bio). The FBPA and FBPAse activities were both measured at 48 °C using a coupled spectrophotometric assay, as described².

FBPase assay. FBP-dependent fructose-6-phosphate formation was measured by coupling the reaction with exogenous phosphoglucose isomerase and glucose-6-phosphate dehydrogenase, and NADPH formation was monitored at 340 nm ($\epsilon_{340 \text{ nm}} \text{ NADPH} = 6,300 \text{ M}^{-1} \text{ cm}^{-1}$). The assay mixture (0.5 ml) consisted of 0.1 M Tris-HCl (pH 7.8), 20 mM MgCl₂, 20 mM dithiothreitol (DTT), 0.5 mM NADP⁺, 0.01–0.3 mM FBP and 1 U each of phosphoglucose isomerase and glucose-6-phosphate dehydrogenase from baker's yeast (Sigma-Aldrich). The reaction was started by the addition of the purified enzyme.

FBPA assay (anabolic direction). Triosephosphate-dependent fructose-6-phosphate formation was measured by coupling the reaction with exogenous

triosephosphate isomerase, phosphoglucose isomerase and glucose-6-phosphate dehydrogenase from baker's yeast (Sigma-Aldrich); NADPH formation was monitored at 340 nm. The assay mixture (0.5 ml) consisted of 0.1 M Tricine-KOH (pH 8.0), 20 mM MgCl₂, 20 mM DTT, 0.5 mM NADP⁺, 4 U of triosephosphate isomerase, 1 U of phosphoglucose isomerase and 1 U of glucose-6-phosphate dehydrogenase. The assay mixture was preincubated for 4 min. After the addition of GA3P (0.04–0.35 mM), the assay mixture was further incubated for 1 min to achieve equilibrium between GA3P and DHAP. The reaction was started by the addition of the purified enzyme.

FBPA assay (catabolic direction). FBP-dependent formation of triosephosphates was measured by coupling the reaction with triosephosphate isomerase from baker's yeast and glycerolphosphate dehydrogenase from rabbit (Sigma-Aldrich), and the oxidation of NADH was monitored at 365 nm ($\epsilon_{365 \text{ nm}} \text{ NADH} = 3,400 \text{ M}^{-1} \text{ cm}^{-1}$). The assay mixture (0.5 ml) consisted of 0.1 M Tricine-KOH (pH 8.0), 20 mM MgCl₂, 20 mM DTT, 0.55 mM NADH, 5 mM FBP, 20 U of triosephosphate isomerase and 2 U of glycerolphosphate dehydrogenase. The reaction was started by the addition of the purified enzyme.

To examine whether the enzyme is active under the crystallization conditions, we measured the FBPA activity (anabolic direction) in an assay mixture (0.5 ml) consisting of 0.1 M Bicine-KOH (pH 9.0), 10% PEG 20,000, 2% dioxane, 5 mM MgCl₂, 0.5 mM NADP⁺, 0.5 mM GA3P, 0.5 mM DHAP, 20 U of triosephosphate isomerase, 5 U of phosphoglucose isomerase and 5 U of glucose-6-phosphate dehydrogenase. To examine whether the FBPA/P–DHAP complex represents a genuine reaction intermediate, we washed several crystals using the crystallization buffer, dissolved them in water and then measured the FBPA activity (anabolic direction).

13. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
14. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
15. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).

Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase

Juan Du^{1*}, Rafael F. Say^{2*}, Wei Lü¹, Georg Fuchs² & Oliver Einsle^{1,3}

Fructose-1,6-bisphosphate (FBP) aldolase/phosphatase is a bifunctional, thermostable enzyme that catalyses two subsequent steps in gluconeogenesis in most archaea and in deeply branching bacterial lineages^{1–3}. It mediates the aldol condensation of heat-labile dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) to FBP⁴, as well as the subsequent, irreversible hydrolysis of the product to yield the stable fructose-6-phosphate (F6P) and inorganic phosphate; no reaction intermediates are released. Here we present a series of structural snapshots of the reaction that reveal a substantial remodelling of the active site through the movement of loop regions that create different catalytic functionalities at the same location. We have solved the three-dimensional structures of FBP aldolase/phosphatase from thermophilic *Thermoproteus neutrophilus*^{5,6} in a ligand-free state as well as in complex with the substrates DHAP and FBP and the product F6P to resolutions up to 1.3 Å. In conjunction with mutagenesis data, this pinpoints the residues required for the two reaction steps and shows that the sequential binding of additional Mg²⁺ cations reversibly facilitates the reaction. FBP aldolase/phosphatase is an ancestral gluconeogenic enzyme optimized for high ambient temperatures^{1,2}, and our work resolves how consecutive structural rearrangements reorganize the catalytic centre of the protein to carry out two canonical reactions in a very non-canonical type of bifunctionality.

Aldolases constitute a distinct group of lyase enzymes that catalyse the stereospecific addition of a nucleophilic donor substrate to an electrophilic acceptor^{7–9}. Class I aldolases are commonly homotetramers¹⁰ and are found in some bacteria, archaea and higher eukaryotes^{11,12}. They activate their donor substrate by forming a Schiff base with the ε-amino group of a conserved lysine residue¹³. Aldolases of class II are found in bacteria and fungi and use divalent metal cations (mostly Zn²⁺, but frequently Fe²⁺ or Co²⁺) to activate the donor nucleophile¹⁴. Although the two classes of aldolases do not show homologies in primary structure and are thus phylogenetically distinct, they are structurally related and belong to the family of (β/α)₈ TIM barrel enzymes^{15–18}. Further members of this family include most other aldolases, such as the homodecameric class IA FBP aldolase^{4,19}, transaldolase, deoxyribose phosphate aldolase, 2-keto-3-deoxy-(6-phospho)-gluconate (KD(P)G) aldolase and 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase²⁰.

Fructose-1,6-bisphosphate aldolase (E.C. 4.1.2.13) catalyses the reversible aldol cleavage of FBP into DHAP and GAP in the Embden–Meyerhof–Parnas pathway²¹ either in glycolysis or gluconeogenesis and in the Calvin–Benson cycle¹. Most recently an enzyme previously described as an archaeal FBP phosphatase in *Thermococcus kodakarensis* and *Sulfolobus tokodaii*²² was shown to be indeed a bifunctional FBP aldolase/phosphatase (FBPAP)². Although this enzyme is physiologically unrelated to any known aldolase, it catalyses the reaction following the class I mechanism²³ involving a lysine Schiff base^{24,25}. The aldolase reaction is a classic case for a fully reversible enzymatic

reaction⁹, but in FBPAP – in spite of a similar mechanism – the *K_m* values for the aldol condensation and the aldol cleavage differ by a factor of 1,000, and the intrinsic phosphatase activity renders the process irreversible². Orthologues of FBPAP are found in most genomes of archaea and early bacterial lineages.

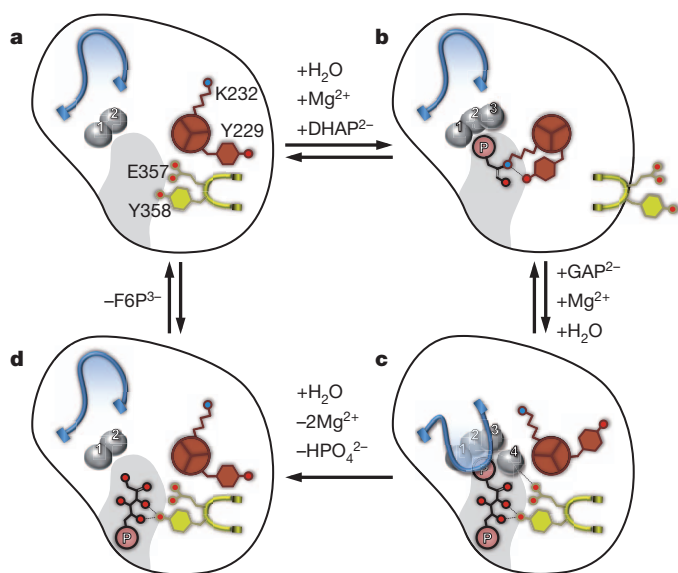
A structure of the enzyme from *S. tokodaii* revealed a novel tertiary structure with a ferredoxin-like fold in the amino (N)-terminal part and similarities to bacterial S-adenosyl methionine decarboxylases in the carboxy (C)-terminal part. The protein formed globular homooctamers and had four Mg²⁺ ions and FBP bound to the active site^{20,22}. To understand the twofold reactivity of FBPAP we have crystallized and characterized the enzyme from the hyperthermophilic crenarchaeon *T. neutrophilus* (*Tn*FBPAP) in a ligand-free state, with the substrates of the two reaction steps, DHAP and FBP, and with the product F6P. *Tn*FBPAP is highly similar to the *S. tokodaii* orthologue, differing predominantly in a C-terminal extension that embraces another protomer of the conserved octamer (Supplementary Fig. 1)²². However, three loop regions surrounding the active site of the enzyme substantially change their conformation when substrate, reaction intermediate or product is bound. The re-orientation of these loops during the catalytic cycle alters the structure and functionality of the active site fundamentally, while keeping the substrate locked in place. Unlike bi- or multifunctional enzymes described previously, FBPAP does not recruit distinct domains for its reactions or connect different active sites by substrate channelling. Instead the enzyme remodels its single active centre in order to bring the amino-acid side chains and cofactors—divalent Mg cations—into place.

In the following, the loop region from residues 220 to 235 will be termed the ‘aldolase loop’, as residues K232 and Y229 are essential for aldolase activity. It is observed in three different conformations that we designate ‘in’, ‘out’ and ‘locked’ (Fig. 1). The largest conformational changes were observed in the loop encompassing residues 89–111, the ‘phosphatase lid’ that serves to fix the intermediate FBP in the binding pocket for the phosphatase reaction. It is seen in two conformations, ‘open’ and ‘closed’. The third flexible loop is the ‘anchor loop’ from residues 353 to 364 that fixes FBP after the aldol condensation and attains an ‘in’ or an ‘out’ conformation. The consecutive binding of additional Mg²⁺ ions to the protein is then key to switching its catalytic functionality.

In the substrate-free structure of *Tn*FBPAP, two Mg²⁺ ions are bound to the protein, with Mg1 coordinated by residues D11, H18 and D52, and Mg2 by residues D52, D53, D132 and D234 from the aldolase loop. Both Mg²⁺ ions show an octahedral coordination environment, completed by three water ligands for Mg1 and by two water ligands for Mg2 (Fig. 2a and Supplementary Figs 2 and 3a). In the structure, a substrate binding cleft is visible close to the ions, and four of the five water ligands are at the surface of the protein. The phosphatase lid is in its ‘open’ conformation, the aldolase loop is ‘out’ and the anchor loop is ‘in’ (Figs 1a and 2a and Supplementary Fig. 3a). A structural analysis of EDTA-treated *Tn*FBPAP shows that Mg1 can

¹Lehrstuhl für Biochemie, Institut für organische Chemie und Biochemie, Albert-Ludwigs-Universität Freiburg, Albertstrasse 21, 79104 Freiburg, Germany. ²Lehrstuhl für Mikrobiologie, Albert-Ludwigs-Universität Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany. ³BIOSS Centre for Biological Signalling Studies, Albert-Ludwigs-Universität Freiburg, Hebelstrasse 25, 79104 Freiburg, Germany.

*These authors contributed equally to this work.



be removed, while Mg2 remains bound to the protein. In a calorimetric titration Mg1 binds with a K_d of 27 μM (Supplementary Fig. 5), but no binding of further magnesium occurs. In this state the enzyme can bind its first substrate, DHAP (but not GAP), with a K_d of 25 μM (Supplementary Fig. 7), thereby creating a further Mg^{2+} binding site. Consequently, the crystal structure with bound DHAP contains three Mg^{2+} , and each of the C1-phosphoryl oxygen atoms coordinates

Figure 1 | Representation of the reaction steps of FBP aldolase/phosphatase. The phosphatase loop is shown in blue, the aldolase loop in red and the anchor loop in yellow. **a**, In the unliganded state the enzyme binds two Mg^{2+} ions and the active site is accessible. **b**, Binding of the substrate dihydroxyacetone phosphate is followed by a third Mg^{2+} ion that redirects the aldolase loop to form a Schiff base with the substrate. The anchor loop is retracted to accommodate this conformational change. **c**, The co-substrate glyceraldehyde-3-phosphate binds and the enzyme catalyses the aldol condensation. Release of the Schiff base leads to a rearrangement of the aldolase loop that creates a binding site for a fourth Mg^{2+} ion. The phosphatase loop changes its conformation drastically closing the active site. **d**, Upon hydrolysis of the phosphate group at C1, the tetranuclear Mg site disassembles and the three loops attain their original conformations.

one of the cations (Figs 1b and 2b and Supplementary Fig. 3b). At both Mg1 and Mg2, two phosphoryl oxygens replace the coordinated water molecules (Supplementary Figs 2 and 3b). In the next step in the reaction cascade, the backbone carbonyl oxygen of residue K232 occupies a remaining, free coordination site at the metal ion. The following major structural rearrangement switches the aldolase loop to the 'in' position and the anchor loop to the 'out' position. The phosphatase lid remains open, but one of its residues, Q95, rearranges to become a ligand to Mg1, likely stabilizing the complex metal site. More importantly, in the 'in' conformation of the aldolase loop K232 is in close proximity to the substrate and is able to form a protonated Schiff base intermediate with DHAP that is visible in the crystal structure (Figs 1b and 2b and Supplementary Fig. 3b).

The class-I aldolase mechanism requires a base that abstracts a proton in the next step^{26,27}, and in *Tn*FBPAP residue Y229 is positioned ideally to fulfil this role. It is part of the aldolase loop and swings

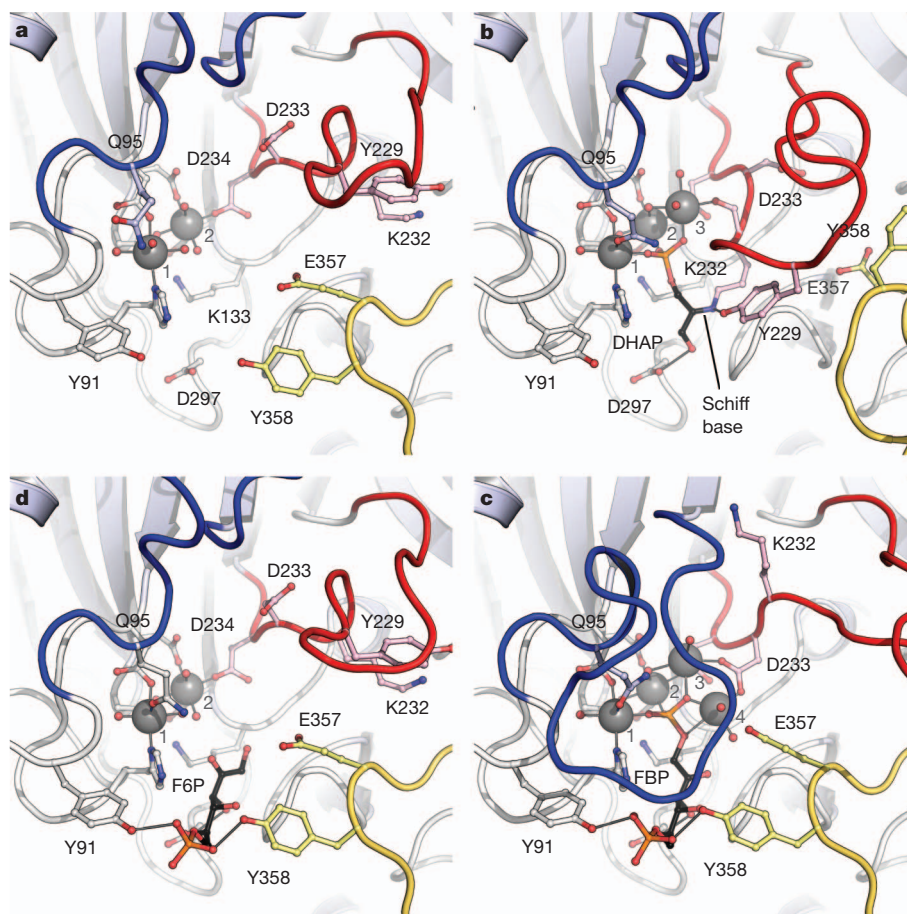


Figure 2 | The successive reaction steps of FBP aldolase/phosphatase in the crystal structures. **a**, The unliganded enzyme; **b**, the complex with dihydroxyacetone phosphate (DHAP); **c**, the complex with fructose-1,6-bisphosphate (FBP); **d**, the product complex with fructose-6-phosphate (F6P).

The reaction is guided by the successive binding of a third and fourth Mg^{2+} ion that orchestrate conformational rearrangements to position the aldolase (red), phosphatase (blue) and anchor loops (yellow) for their individual tasks.

into the active site when K232 forms the Schiff base. The distinct and isolated role of K232 and Y229 is emphasized by the fact that variants in these positions show only slightly reduced (Y229F) or even enhanced (K232R) phosphatase activity, whereas aldolase activity is completely abolished (Fig. 3 and Supplementary Table 2). The structure of the Y229F variant of FBPAP shows binding of DHAP with two Mg^{2+} ions, but the formation of the Schiff base is not observed. After its deprotonation, Y229 abstracts the *pro-S* proton at C3 of DHAP. This is ascertained by residue D297 that forms a short (2.6 Å) hydrogen bond with the hydroxyl group at C3 so that the *pro-S* H atom is fixed to face Y229 (Supplementary Fig. 3b). Proton abstraction results in a tautomeric rearrangement to yield an enamine intermediate. For the progression of the aldol condensation, the second substrate, GAP, then needs to bind to the active site. In the structure, the binding site for GAP is occupied by a second molecule of DHAP, giving a clear indication for the binding mode of the substrate, while being unreactive for the aldol condensation (Supplementary Fig. 3b).

C–C bond formation between the DHAP enamine and GAP yields the product Schiff base (Fig. 4) that is released to trigger the next conformational rearrangement in preparation for the phosphatase step. As the aldolase loop flips outward, the anchor loop changes back to the ‘in’ conformation, and one of its key residues, Y358, forms hydrogen bonds to the C4 OH group and the C6 phosphate of FBP (Figs 1c and 2c and Supplementary Fig. 3c). The phosphatase lid switches to ‘closed’, preventing access to the active site cavity. Concomitantly, the aldolase loop attains its ‘locked’ conformation, wherein K232 forms a hydrogen bond to the backbone carbonyl group of P111 at the base of the phosphatase loop (Supplementary Fig. 4). In effect, the position of the neighbouring D233 inverts, and its β -carboxy group becomes key to the creation of an additional binding site for Mg_4 . Two water ligands at Mg_4 are hydrogen-bonded to the side chain of E357 in the anchor loop and the four Mg^{2+} cations of this functional state are tightly grouped around the C1 phosphate of FBP (Supplementary Fig. 3b). The crystal structure, at a resolution of 1.3 Å, indeed shows a clear geometric distortion around the phosphorus atom, resulting in an O–P–O angle of approximately 90° for the two oxygen atoms coordinating Mg_4 (Fig. 2c). Note that before the binding of Mg_4 the three other Mg^{2+} ions are bound on the same side of the

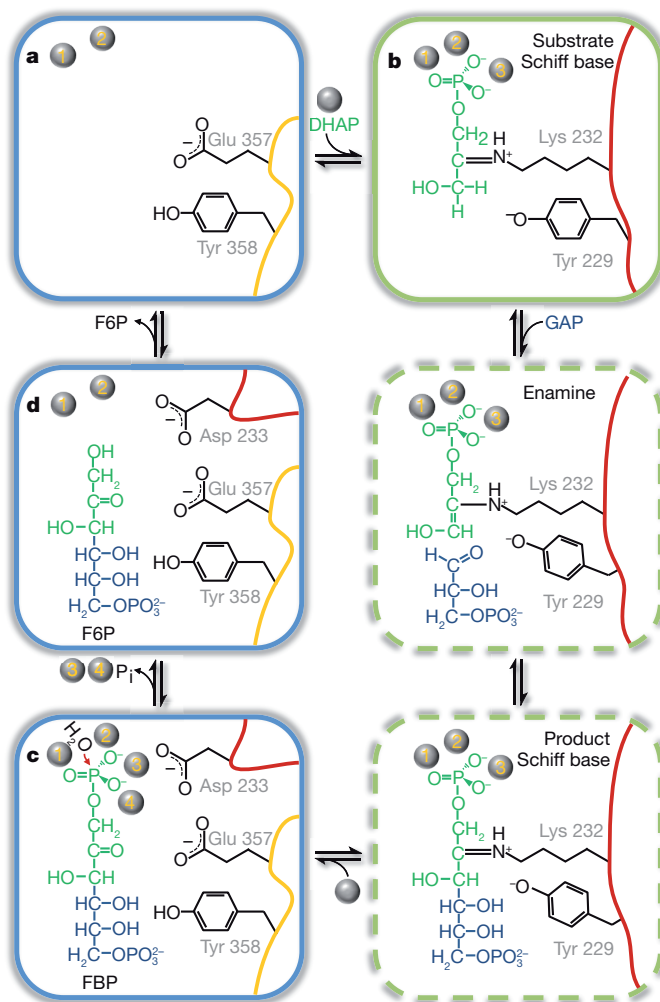


Figure 4 | Proposed reaction mechanism of FBP aldolase/phosphatase. The steps of the aldol condensation reaction are boxed in green, those of the phosphatase reaction in blue. Crystal structures are available for the intermediates in solid boxed, whereby the lettering follows Figs 1 and 2. The aldolase loop is indicated by a red line, the anchor loop by a yellow line.

phosphate group (Fig. 2b), thereby counteracting the formation of the trigonal bipyramidal transition state of phosphate ester hydrolysis. Variants of the anchor loop residues E357Q and Y358F exhibit unaltered aldolase activity, whereas phosphatase activity is fully abolished (Fig. 3). As described for the *S. tokodaii* enzyme, the structure allows the identification of a water molecule coordinated by Mg_2 and Mg_3 that conducts an in-line S_N2 nucleophilic attack onto the phosphorus atom (Supplementary Fig. 2). This leads to an inversion of the ligand environment and the cleavage of the phosphoester bond to the sugar substrate.

At this point, the particular mechanistic intricacy of *ThFBPAP* becomes apparent. After binding of DHAP, the aldolase and phosphatase reaction steps were supported by the subsequent addition of tightly coordinated Mg^{2+} ions. The metal triad or tetrad was organized and stabilized by the phosphate group at its centre, but as this phosphate inverts its geometry upon hydrolysis it no longer fits the arrangement of the surrounding cations. The significant free enthalpy of the hydrolysis reaction provides the driving force to overcome the stability of the complex and disassemble the centre. These changes lead back to the structure of the complex with the product F6P in the open-chain form. This structure contains only two Mg^{2+} ions and shows all three loops in the conformations observed in the ligand-free form of the enzyme (Figs 1d and 2d and Supplementary Fig. 3d). Inorganic phosphate itself was not found to bind in crystal soaking experiments.

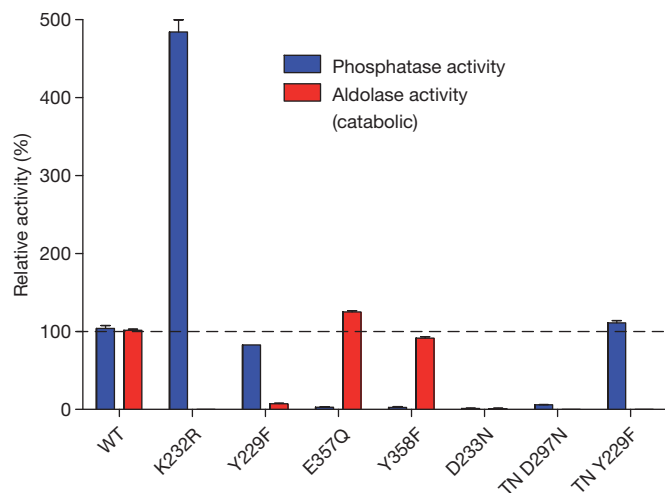


Figure 3 | Catalytic activity of distinct variants of FBP aldolase/phosphatase. The two activities of the enzyme are tightly linked to distinct amino-acid residues and can be inactivated separately. Mutagenesis of K232 or Y229 on the aldolase loop abolishes aldolase activity while phosphatase activity is only slightly reduced (Y229F) or even enhanced (K232R). In contrast, E357 and Y358 on the anchor loop are essential for phosphatase activity, but alterations here (E357Q, Y358F) do not affect aldolase activity. A D233N mutant protein was impaired in both aldolase and phosphatase activity. WT, wild type.

TnFBPAP catalyses a multi-step reaction by remodelling its active site according to the respective catalytic requirements (Fig. 4). At first glance, this seems more elaborate than to combine different enzymatic modules into a multi-enzyme complex as seen, for example, in pyruvate dehydrogenase^{28,29}. Yet TnFBPAP likely represents an ancestral gluconeogenic enzyme and the functional optimizations observed are an adaptation to high-temperature environments, where the instability of the substrates DHAP and GAP presents a serious problem². Large conformational changes and the coupling to the exergonic phosphate hydrolysis allow the enzyme to render the aldolase reaction irreversible to assure that the cellular levels of the sensitive triose phosphates can be kept low. Besides its evolutionary impact, this enzyme sets an elucidatory example for how consecutive dramatic conformational changes can reorganize an active centre to perform two drastically different catalytic steps in a highly controlled and ordered sequence.

METHODS SUMMARY

C-terminally His₆-tagged TnFBPAP (Tneu_0133) was heterologously produced in *Escherichia coli* as described previously². Crystals were grown at 20 °C by the hanging-drop vapour-diffusion method. Two microlitres of protein solution were mixed with 2 µl of a reservoir solution containing 8% (w/v) of polyethylene glycol 3350 and 0.1 M HEPES/NaOH at pH 7.0–8.0. The drops were equilibrated against reservoir solution without added protein. The crystals were briefly soaked in 15% (v/v) of 2R,3R-butane diol for cryoprotection before being plunged into liquid nitrogen. For the preparation of complexes with the compounds dihydroxyacetone phosphate, fructose-1,6-bisphosphate and fructose-6-phosphate, the native crystals were soaked in 15% (v/v) of 2R,3R-butane diol and 100 mM of the respective substrate for 20 min before being treated for cryo-protection as described above. Diffraction data were collected at beamline X06SA at the Swiss Light Source (Villigen, Switzerland) at an X-ray wavelength of 1.0 Å. The native structure was refined to $R_{\text{cryst}} = 0.165$ and $R_{\text{free}} = 0.188$ at a resolution of 1.52 Å. The crystal structures in complex with DHAP, FBP and F6P were solved by molecular replacement with the native structure as the search model (Supplementary Table 1). Activity assays for the *Cenarchaeum symbiosum* and *T. neutrophilus* enzymes were performed at 40 °C and 48 °C, respectively, using a coupled spectrophotometric assay as described previously².

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 March; accepted 13 August 2011.

Published online 9 October 2011.

1. Berg, I. A. *et al.* Autotrophic carbon fixation in archaea. *Nature Rev. Microbiol.* **8**, 447–460 (2010).
2. Say, R. F. & Fuchs, G. Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464**, 1077–1081 (2010).
3. Stetter, K. O. Hyperthermophiles in the history of life. *Phil. Trans. R. Soc. B* **361**, 1837–1842 (2006).
4. Siebers, B. *et al.* Archaeal fructose-1,6-bisphosphate aldolases constitute a new family of archaeal type class I aldolase. *J. Biol. Chem.* **276**, 28710–28718 (2001).
5. Messner, P., Pum, D., Sara, M., Stetter, K. O. & Sleytr, U. B. Ultrastructure of the cell envelope of the archaeobacteria *Thermoproteus tenax* and *Thermoproteus neutrophilus*. *J. Bacteriol.* **166**, 1046–1054 (1986).
6. Zillig, W. *et al.* The phylogenetic relations of DNA-dependent RNA polymerases of archaeobacteria, eukaryotes, and eubacteria. *Can. J. Microbiol.* **35**, 73–80 (1989).
7. Alefounder, P. R., Baldwin, S. A., Perham, R. N. & Short, N. J. Cloning, sequence analysis and over-expression of the gene for the class II fructose 1,6-bisphosphate aldolase of *Escherichia coli*. *Biochem. J.* **257**, 529–534 (1989).
8. Fothergill-Gilmore, L. A. & Michels, P. A. Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* **59**, 105–235 (1993).
9. Rutter, W. J. Evolution of aldolase. *Fed. Proc.* **23**, 1248–1257 (1964).

10. Lorentzen, E., Siebers, B., Hensel, R. & Pohl, E. Structure, function and evolution of the archaeal class I fructose-1,6-bisphosphate aldolase. *Biochem. Soc. Trans.* **32**, 259–263 (2004).
11. Penhoet, E., Kochman, M., Valentine, R. & Rutter, W. J. The subunit structure of mammalian fructose diphosphate aldolase. *Biochemistry* **6**, 2940–2949 (1967).
12. Tolan, D. R., Niclas, J., Bruce, B. D. & Lebo, R. V. Evolutionary implications of the human aldolase-A, -B, -C, and -pseudogene chromosome locations. *Am. J. Hum. Genet.* **41**, 907–924 (1987).
13. Lebherz, H. G. & Rutter, W. J. Distribution of fructose diphosphate aldolase variants in biological systems. *Biochemistry* **8**, 109–121 (1969).
14. Cooper, S. J. *et al.* The crystal structure of a class II fructose-1,6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold. *Structure* **4**, 1303–1315 (1996).
15. Lorentzen, E. *et al.* Crystal structure of an archaeal class I aldolase and the evolution of (β α)₈ barrel proteins. *J. Biol. Chem.* **278**, 47253–47260 (2003).
16. Hester, G. *et al.* The crystal structure of fructose-1,6-bisphosphate aldolase from *Drosophila melanogaster* at 2.5 Å resolution. *FEBS Lett.* **292**, 237–242 (1991).
17. Gambin, S. J. *et al.* The crystal structure of human muscle aldolase at 3.0 Å resolution. *FEBS Lett.* **262**, 282–286 (1990).
18. Blom, N. S., Tetreault, S., Coulombe, R. & Sygusch, J. Novel active site in *Escherichia coli* fructose-1,6-bisphosphate aldolase. *Nature Struct. Biol.* **3**, 856–862 (1996).
19. Imanaka, H., Fukui, T., Atomi, H. & Imanaka, T. Gene cloning and characterization of fructose-1,6-bisphosphate aldolase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Biosci. Bioeng.* **94**, 237–243 (2002).
20. Andreeva, A. *et al.* Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **36**, 419–425 (2008).
21. Ronimus, R. S. & Morgan, H. W. Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**, 199–221 (2003).
22. Nishimasu, H., Fushinobu, S., Shoun, H. & Wakagi, T. The first crystal structure of the novel class of fructose-1,6-bisphosphatase present in thermophilic archaea. *Structure* **12**, 949–959 (2004).
23. St-Jean, M., Blonski, C. & Sygusch, J. Charge stabilization and entropy reduction of central lysine residues in fructose-bisphosphate aldolase. *Biochemistry* **48**, 4528–4537 (2009).
24. Grazi, E., Rowley, P. T., Cheng, T., Tchola, O. & Horecker, B. L. The mechanism of action of aldolases. III. Schiff base formation with lysine. *Biochem. Biophys. Res. Commun.* **9**, 38–43 (1962).
25. Rose, I. A. & Rieder, S. V. Studies on the mechanism on the aldolase reaction; isotope exchange reactions of muscle and yeast aldolase. *J. Biol. Chem.* **231**, 315–329 (1958).
26. Lai, C. Y., Tchola, O., Cheng, T. & Horecker, B. L. The mechanism of action of aldolases. 8. The number of combining sites in fructose diphosphate aldolase. *J. Biol. Chem.* **240**, 1347–1350 (1965).
27. St-Jean, M., Lafrance-Vanasse, J., Liotard, B. & Sygusch, J. High resolution reaction intermediates of rabbit muscle fructose-1,6-bisphosphate aldolase: substrate cleavage and induced fit. *J. Biol. Chem.* **280**, 27262–27270 (2005).
28. Ciszak, E. M., Korotchkina, L. G., Dominiak, P. M., Sidhu, S. & Patel, M. S. Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase. *J. Biol. Chem.* **278**, 21240–21246 (2003).
29. Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu. Rev. Biochem.* **69**, 961–1004 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Deutsche Forschungsgemeinschaft (grant Ei520/3-2 to O.E., Fu118/15-4 and 15-5 to G.F.). Diffraction data were collected at beam lines X06SA and X06DA at the Swiss Light Source (Villigen, Switzerland). The authors thank the beam line staff for assistance during data collection.

Author Contributions J.D. and W.L. crystallized the protein and collected diffraction data, J.D., W.L. and O.E. built and refined the structural models, R.F.S. created and analysed the variant proteins, J.D., R.F.S., W.L., G.F. and O.E. designed the experiments, O.E. wrote the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures are deposited in Protein Data Bank under accession numbers 3T2B (ligand-free), 3T2C (DHAP-bound), 3T2D (FBP-bound), 3T2E (F6P-bound), 3T2F (EDTA-soak with DHAP) and 3T2G (Y229F variant with DHAP). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.E. (einsle@biochemie.uni-freiburg.de).

METHODS

Cloning, expression and purification. C-terminally His₆-tagged *TnFBPAP* (Tneu_0133) was heterologously produced in *E. coli* as described previously². Frozen *E. coli* cells (30 g wet mass) were suspended in 30 ml re-suspension buffer (20 mM Tris/HCl pH 7.8, 20 mM MgCl₂, 0.5 mM phenylmethanesulphonyl fluoride) containing 0.1 mg ml⁻¹ DNase I and passed twice through a French pressure cell at 137 MPa. The cell lysate was centrifuged at 150,000g for 1 h (4 °C). The cell extract (10 ml) was loaded onto a 1 ml Ni²⁺-chelating Sepharose affinity column (GE Healthcare) equilibrated with running buffer (20 mM Tris/HCl pH 7.8, 50 mM MgCl₂ and 200 mM NaCl) at a flow rate of 1 ml min⁻¹. The column was washed with running buffer containing 70 mM imidazole and developed with a linear gradient of imidazole. Recombinant *TnFBPAP* eluted at a concentration of 275 mM imidazole. Active fractions were pooled and the buffer was exchanged for 50 mM 3-(N-morpholino)propanesulphonate (MOPS) buffer at pH 8.0 with 10 mM MgCl₂ using a PD-10 column (GE Healthcare). For subsequent heat precipitation, the enzyme was incubated for 30 min at 80 °C, cooled on ice for 15 min and centrifuged at 17,000g for 30 min (4 °C). For crystallization, the supernatant with purified protein was concentrated by ultrafiltration to a final concentration of 7 mg ml⁻¹.

Crystallization. Crystals were grown at 20 °C by the hanging-drop vapour-diffusion method. Two microlitres of protein solution were mixed with 2 µl of a reservoir solution containing 8% (w/v) of polyethylene glycol 3350 and 0.1 M HEPES/NaOH at pH 7.0–8.0. The drops were equilibrated against reservoir solution without added protein. Native crystals appeared after 1 h and reached their maximum size after approximately 3 days. The crystals were briefly soaked in 15% (v/v) of 2R,3R-butane diol for cryoprotection before being plunged into liquid nitrogen. For the preparation of complexes with the compounds dihydroxyacetone phosphate, fructose-1,6-bisphosphate and fructose-6-phosphate, the native crystals were soaked in 15% (v/v) of 2R,3R-butane diol and 100 mM of the respective substrate for 20 min before being treated for cryo-protection as described above.

Data collection and structure determination. Diffraction data were collected at beamline X06SA at the Swiss Light Source (Villigen, Switzerland) at an X-ray wavelength of 1.0 Å. The crystals of *TnFBPase* belonged to the tetragonal space group *I*422 with one monomer per asymmetric unit. Data were indexed and integrated using MOSFLM³⁰ and scaled with sCALA³¹. Structure solution was carried out by molecular replacement using the program MOLREP³² and the structure of the homologous enzyme from *S. tokodaii* (PDB-ID 1UMG) as the initial search model. Refinement was performed with REFMAC5 (ref. 33) and model building was performed using coot³⁴. The native structure was refined to $R_{\text{cryst}} = 0.165$ and $R_{\text{free}} = 0.188$ (ref. 35) at a resolution of 1.52 Å. The crystal

structures in complex with DHAP, FBP and F6P were solved by molecular replacement with the native structure as the search model (Supplementary Table 1).

Site-directed mutagenesis. Mutagenesis was performed on the synthetic *C. symbiosum* gene¹. The first mutations were introduced into the expression vector pT7-7 (ref. 36) carrying the gene (FBP_C.symb-X-pT7-7 (ref. 2)) by reverse PCR using a single mutagenic oligonucleotide⁹. PCR conditions were as follows: 25 cycles of 20 s denaturation at 98 °C, 20 s primer annealing, and elongation at 72 °C using Phusion DNA polymerase (New England Biolabs). The PCR products were incubated at 37 °C with 20 U of *DpnI* for 3 h to digest the methylated template plasmid. After amplification of the plasmid in *E. coli* DH5 α , the mutation was confirmed by sequencing. Competent *E. coli* BL21 (DE3) Rosetta2 cells (Novagen) were transformed with the corresponding plasmid and grown at 37 °C in 1-l flasks with self-inducing medium³⁷ containing 100 µg ml⁻¹ ampicillin and 34 µg ml⁻¹ chloramphenicol. After 5–6 h at 37 °C, the temperature was lowered to 20 °C and the culture was grown for 14 h at 20 °C. Cells were collected by centrifugation and stored in liquid nitrogen until further processing.

Production of FBP aldolase/phosphatase mutants from *C. symbiosum*. Preparation of cell extract from 3–6 g wet mass of frozen *E. coli* cells and affinity chromatography were performed following the procedure described above for the *T. neutrophilus* protein.

FBP aldolase/phosphatase enzyme assays. Activity assays for the *C. symbiosum* and *T. neutrophilus* enzymes were performed at 40 °C and 48 °C, respectively, using a coupled spectrophotometric assay as described previously².

30. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* No. 26, (1992).
31. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
32. Vagin, A. A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
33. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Brünger, A. T. Assessment of phase accuracy by cross validation - the free R-value - methods and applications. *Acta Crystallogr. D* **49**, 24–36 (1993).
36. Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl Acad. Sci. USA* **82**, 1074–1078 (1985).
37. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

Efficient quantum computing using coherent photon conversion

N. K. Langford^{1,2,3}, S. Ramelow^{1,2}, R. Prevedel^{1,4}, W. J. Munro^{5,6}, G. J. Milburn^{7,1} & A. Zeilinger^{1,2}

Single photons are excellent quantum information carriers: they were used in the earliest demonstrations of entanglement¹ and in the production of the highest-quality entanglement reported so far^{2,3}. However, current schemes for preparing, processing and measuring them are inefficient. For example, down-conversion provides heralded, but randomly timed, single photons⁴, and linear optics gates are inherently probabilistic⁵. Here we introduce a deterministic process—coherent photon conversion (CPC)—that provides a new way to generate and process complex, multiqumata states for photonic quantum information applications. The technique uses classically pumped nonlinearities to induce coherent oscillations between orthogonal states of multiple quantum excitations. One example of CPC, based on a pumped four-wave-mixing interaction, is shown to yield a single, versatile process that provides a full set of photonic quantum processing tools. This set satisfies the DiVincenzo criteria for a scalable quantum computing architecture⁶, including deterministic multiqubit entanglement gates (based on a novel form of photon–photon interaction), high-quality heralded single- and multiphoton states free from higher-order imperfections, and robust, high-efficiency detection. It can also be used to produce heralded multiphoton entanglement, create optically switchable quantum circuits and implement an improved form of down-conversion with reduced higher-order effects. Such tools are valuable building blocks for many quantum-enabled technologies. Finally, using photonic crystal fibres we experimentally demonstrate quantum correlations arising from a four-colour nonlinear process suitable for CPC and use these measurements to study the feasibility of reaching the deterministic regime with current technology^{4,7}. Our scheme, which is based on interacting bosonic fields, is not restricted to optical systems but could also be implemented in optomechanical, electromechanical and superconducting systems^{8–12} with extremely strong intrinsic nonlinearities. Furthermore, exploiting higher-order nonlinearities with multiple pump fields yields a mechanism for multiparty mediation of the complex, coherent dynamics.

One of the key challenges for photonic quantum information processing is to induce strong, deterministic interactions between individual photons, which cannot be done using standard linear optical components. The scheme proposed for linear optics quantum computing in refs 5, 13 avoided this problem by using the inherent nonlinearity of photodetection and non-classical interference to induce effective nonlinear photon interactions non-deterministically. Alternatively, in the one-way picture of quantum computing, the required nonlinearities are replaced by offline probabilistic preparation of special entangled states followed by detection and feed-forward^{14–16}.

Nonlinear optics quantum computing (NLOQC) takes a different approach, by directly using intrinsic nonlinearities to implement multiphoton interactions. NLOQC schemes using different types of optical nonlinearity, including cross-Kerr coupling^{17,18} and two-photon absorption¹⁹, have been put forward. Since those proposals were made, more-complete multimode analyses of the cross-Kerr NLOQC schemes

have suggested that they cannot in fact produce phase shifts large enough for NLOQC because of spectral correlations created between the interacting fields²⁰. Other work, however, shows that these difficulties can be circumvented in the related case of strong, second-order nonlinear ($\chi^{(2)}$) interactions by carefully engineering the phase-matching conditions²¹.

Coherent photon conversion is an alternative nonlinear approach that uses coherent oscillations between different multi-excitation states. The underlying process is a nonlinear interaction between m bosonic modes that coherently converts single excitations in some of the modes (depending on the precise form of the interaction) into single excitations in the remaining modes. A key principle of CPC is that this basic nonlinearity can in turn be generated by pumping some modes of a higher-order nonlinearity with strong classical fields. This induces an effective coupling between the quantum modes that can be tuned and enhanced by the classical pumps, even to the point where the effective interaction is stronger than naturally occurring couplings of the same form. A similar effect is achieved in photon-pair sources based on four-wave mixing in photonic crystal fibres. Such systems have produced some of the highest-brightness photon-pair sources with very low pump powers⁷, and precise dispersion engineering and fibre-structuring technologies have allowed optimization of these sources to produce ultrabright, high-purity, heralded single photons⁴.

To illustrate the potential of CPC, here we focus on a novel case that has very interesting properties for quantum optics applications and which is based on the following standard four-wave-mixing interaction involving four distinct frequency modes (a , b , c and d):

$$H = \gamma ab^\dagger c^\dagger d + \gamma^* a^\dagger bcd^\dagger \quad (1)$$

Here the coupling strength, γ , arises from the third-order ($\chi^{(3)}$) nonlinearity, an asterisk denotes complex conjugate and a dagger denotes operator adjoint. Pumping mode d with a bright classical beam with electric field amplitude E yields the effective second-order interaction

$$\tilde{H} = \tilde{\gamma} ab^\dagger c^\dagger + \tilde{\gamma}^* a^\dagger b c \quad (2)$$

where $\tilde{\gamma} \propto \gamma E$. This now resembles a standard three-wave-mixing Hamiltonian with an enhanced, tunable, nonlinear coupling.

The key to understanding how this CPC process works, and its potential, is to note that an input Fock state, $|n_a n_b n_c\rangle$, will evolve within a well-defined, restricted Hilbert space, $\{\tilde{H}^j |n_a n_b n_c\rangle | \forall \text{ integers } j\}$, of dimension $n_a + \min(n_b, n_c) + 1$. Consequently, it will exhibit the collapses and revivals of individual population elements that are characteristic of coherent quantum processes. Most importantly for our scheme, for the two-dimensional subspace $\{|100\rangle, |011\rangle\}$ ($\tilde{H}|100\rangle \propto |011\rangle$ and $\tilde{H}^2|100\rangle \propto |100\rangle$), the induced coupling drives Rabi-like oscillations between the two basis states. Given the input state $|100\rangle$, the output is

$$|\psi(t)\rangle = \cos(\Gamma t)|100\rangle + i \frac{\tilde{\gamma}}{|\tilde{\gamma}|} \sin(\Gamma t)|011\rangle \quad (3)$$

where $\Gamma = |\tilde{\gamma}|/\hbar$ (\hbar being Planck's constant divided by 2π).

¹Vienna Center for Quantum Science and Technology, Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria. ²Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria. ³Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK. ⁴Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. ⁵National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. ⁶NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan. ⁷Centre for Engineered Quantum Systems, University of Queensland, St Lucia, 4072 Queensland, Australia.

Notably, standard single-photon up-conversion, a special case of CPC (indeed the simplest case), is one of the small subset of CPC processes with purely classical analogues. If a classical input field is used, then, provided that this ‘pump’ field remains undepleted throughout, the input field will undergo complete coherent oscillations between the two frequency modes. By contrast, if a classical input is used in the above example, then no coherent oscillations will be observed: the output is the well-known two-mode squeezed state of a parametric down-conversion source. In other words, in most cases a key element of CPC operation is the use of quantized inputs.

The basic conditions for a viable implementation of quantum computing are the DiVincenzo criteria⁶. The major unresolved challenges for photonic quantum information processing are good multiphoton sources, reliable multiqubit interactions and robust, high-efficiency single-photon detection. We show here that CPC provides tools to solve all three of these issues (Figs 1 and 2), all derived from a single process just by choosing different interaction strengths.

Figure 1a shows how CPC directly implements a two-qubit controlled-Z gate between the photons in the two modes *b* and *c*. The key insight is that CPC, like any coherent process that cycles between two orthogonal states, gives rise to geometric (Berry’s) phase effects^{22–25}. Therefore, for $t = \pi/\Gamma$, an input state $|011\rangle$ will undergo a full oscillation and undergo a phase shift of π , giving the final state $-|011\rangle$. This directly implements a maximally entangling, controlled-phase gate with 100% efficiency. We note that this geometric phase is truly non-classical and has no equivalent with classical input states. This controlled-Z gate can also be switched very fast optically (using the classical pump²⁵), allowing the fast, real-time ‘rewiring’ of optical quantum circuits. This may have application in various adaptive quantum algorithms, such as quantum phase estimation, and might be particularly useful in waveguide and integrated-optics architectures.

If the input state undergoes half an oscillation ($t = \pi/2\Gamma$), a single photon can be converted coherently and deterministically (with 100% efficiency) into two single photons—the process is a deterministic photon doubler²⁶ (or, in reverse, a deterministic two-photon absorber). Figure 1b illustrates one method for implementing a scalable photon

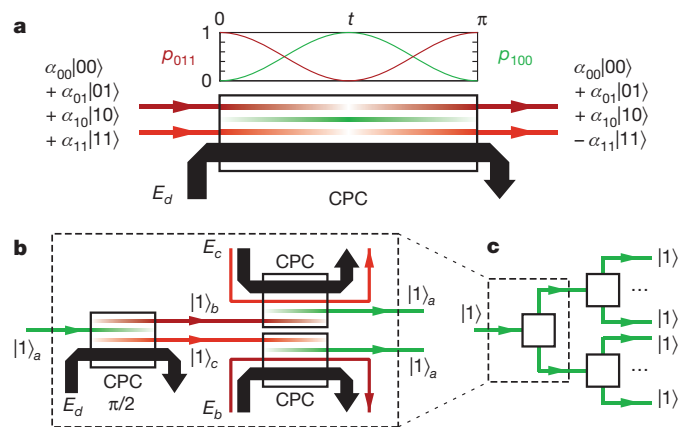


Figure 1 | Satisfying the DiVincenzo criteria with CPC. **a**, Deterministic controlled-phase gate. A ‘ π ’ CPC interaction ($t = \pi/\Gamma$) is an effective photon–photon interaction that implements an entangling controlled-Z gate between two logical states (for example polarization or spatial encoding) of photons with different frequencies. The input state, $|\psi(0)\rangle = \sum_{ij} \alpha_{ij}|ij\rangle \equiv \sum_{ij} \alpha_{ij}|0ij\rangle$ is defined according to so-called ‘single-rail’ logic, where $|ij\rangle$ denotes a state with *i* photons in mode *b* and *j* photons in mode *c*. E_d denotes a bright classical pump field in mode *d*. **b**, Scalable element for deterministic photon doubling. A ‘ $\pi/2$ ’ CPC interaction ($t = \pi/2\Gamma$) can be used both to convert any single-photon source into a good source of multiphoton states and to perform high-efficiency, low-noise detection at any wavelength. E_j again denotes a bright classical pump field in mode *j*, and $|1\rangle_j$ denotes a single-photon Fock state in mode *j*. **c**, Deterministic photon-doubling cascade. The scalable photon doubler from **b** (represented by the symbol in the dashed box) can be directly chained with others to create a deterministic cascade for either multiphoton state preparation or detection enhancement.

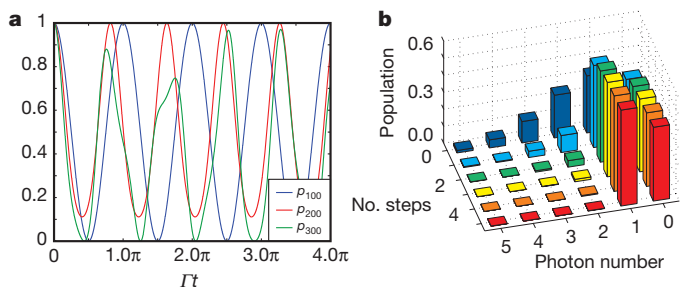


Figure 2 | Heralded single-photon source. **a**, Evolution of $|n_a, 00\rangle$ populations under the CPC interaction for $n_a = 1, 2$ and 3 . This can be used to determine the optimal interaction length for preparing Fock states of different photon number. For example, like the controlled-Z gate, the one-photon Fock state preparation requires an interaction strength (Γt) corresponding to one full cycle of the state $|100\rangle$ (compared with one full cycle of $|011\rangle$ in the case of the controlled-Z gate). **b**, Number-state populations after each filtering step for $t = \pi/\Gamma$, giving $|1\rangle_a$. Combined with a single photon-doubling step and given a weak coherent input state with $|\alpha|^2 = 1.5$, in only five steps this scheme gives heralded single photons with high efficiency ($\sim 56\%$) and minimal higher-order terms ($< 0.3\%$).

doubler, allowing multiple doublers to be chained together to create an arbitrary number of photons (Supplementary Information). This efficient photon-doubling cascade (Fig. 1c) can be used to create a high-quality, scalable source of multiphoton states from any source of genuine single photons (on-demand or heralded). We note that the photon doubler can also be used in conjunction with existing methods to create arbitrary, heralded (also non-locally prepared) Bell-type two-photon and Greenberger–Horne–Zeilinger-type three-photon entanglement, and that these tools can directly implement the encoding step for a simple nine-qubit error correction scheme²⁷ (Supplementary Information).

The same photon-doubling cascade (Fig. 1b) can also be used to perform high-efficiency, low-noise detection with real-world noisy, inefficient detectors by preceding detection with a ‘photon avalanche’ (Supplementary Information). Moreover, this technique can produce marked improvements in detector characteristics, even when the photon doubling efficiencies are less than 100%.

Finally, CPC can also be used to create a high-fidelity source of heralded single photons that could be used to seed the efficient photon-doubling cascade described above. As noted previously, higher-order input states, $|n_a, 00\rangle$, will evolve within a restricted, $(n_a + 1)$ -dimensional Hilbert space. As in the qubit case, this leads to coherent oscillations of population (see Fig. 2a for $n_a = 1, 2$ and 3), but their complexity increases rapidly as n_a increases, because the evolution is governed by an increasingly complicated distribution of eigenfrequencies (see, for example, the $n_a = 3$ case in Fig. 2a, and see Supplementary Information for details). As more competing frequencies come into play, for higher orders these oscillations are characterized by collapses and revivals in the input state population at often irregular times. Remarkably, these frequencies are incommensurate with the frequencies from other orders, so the revivals occur at different times for different input states (Fig. 2a).

We therefore consider an input state in mode *a* that is a superposition (or mixture) of states with different values of n_a , for example $|\psi(0)\rangle_a = |\alpha\rangle_a$ (the latter a ‘classical’ coherent state). After one complete oscillation of the $|100\rangle$ term (that is, after $t = \pi/\Gamma$, as for the controlled-Z gate), all other terms will, with non-zero probability, have converted into states with photons in modes *b* and *c*, which can be rejected using spectral filtering. Applying this process repeatedly will suppress all contributions from other orders, leaving only the $|1\rangle_a$ state with a finite probability (Fig. 2b). (By detecting the dump port of the filtering step with high efficiency and rejecting trials that lead to detection events in these arms, this acts like a pure filter for Fock states²⁸.) By combining this process with a single coherent photon-doubling step, it becomes a heralded single-photon source.

At present, spontaneous parametric down-conversion and spontaneous four-wave mixing provide the best available sources of heralded single photons, but the performance and achievable rates of these sources are intrinsically limited by the effects of higher-order photon-number terms²⁹. By contrast, given a simple, weak coherent input state with $|\alpha|^2 = 1.5$, for example, in only five steps our scheme provides heralded single photons with production efficiencies of $\sim 56\%$ and virtually no higher-order photon-number terms ($< 0.3\%$) (Fig. 2b).

Using similar principles, CPC can also be used to create probabilistically other small Fock states with high fidelity (for example $|200\rangle$) and to implement an improved form of down-conversion that can provide substantially higher pair-emission probabilities with much higher 'heralded' state fidelity than a standard down-conversion source with comparable emission rates (Supplementary Information).

From a practical perspective, there are several advantages to using a pumped $\chi^{(3)}$ interaction to produce an effective $\chi^{(2)}$ nonlinearity. First, although the $\chi^{(3)}$ nonlinearity for a material is normally much weaker than the corresponding $\chi^{(2)}$ nonlinearity (often by many orders of magnitude), the enhancement by the classical field can, for a sufficiently strong pump, result in an effective $\chi^{(2)}$ nonlinear interaction that is stronger than the available natural $\chi^{(2)}$ interaction. Also, materials with inversion symmetry have no $\chi^{(2)}$ nonlinearity ($\chi^{(2)} = 0$), whereas all materials possess a $\chi^{(3)}$ nonlinearity. For example, amorphous glasses, for which there are highly advanced processing technologies, have a $\chi^{(3)}$ nonlinearity but no $\chi^{(2)}$ nonlinearity. Using the classical pump creates a quadratic nonlinearity that is tunable and no longer constrained by fixed material properties. Finally, and perhaps most importantly, conservation of energy allows the four-mode interaction to take place between nearly degenerate frequency modes, which makes CPC compatible with standard telecommunication-band, fibre-based implementations, unlike standard $\chi^{(2)}$ interactions, in which the pump frequency must, by energy conservation, be the sum of the frequencies of the other two photons.

In our proof-of-principle experiments, we study a four-colour interaction in a photonic crystal fibre (PCF) that is a potential candidate for implementing CPC, and investigate the feasibility of reaching the deterministic regime. Specifically, we demonstrate the principle of creating a tunable $\chi^{(2)}$ nonlinearity from a pumped $\chi^{(3)}$ interaction, using a standard commercial, polarization-maintaining PCF, pumped with a 532-nm pulsed laser (Fig. 3). We then use weak coherent states from a 710-nm diode laser in mode a and characterize the feasible interaction strength using the resulting double-pumped correlated-pair source, with output photons with respective wavelengths of 504 and 766 nm.

Figure 4a shows the signal observed for an average power of ~ 90 mW in the 532-nm pulses and an average power of ~ 8.5 mW from the 710-nm diode laser. The diode laser delivers around 10^5 photons during each 532-nm pulse (that is, an effective power of ~ 4.8 μ W), more than four orders of magnitude fewer than the pulsed pump. The background signals measured when only one beam is present show that the observed peak is a combined effect of both input beams. The full time trace allows us to correct very precisely for the periodic background and to isolate the signal that arises from the target four-colour nonlinear interaction that underlies our four-mode CPC process (Fig. 4b).

Figure 4c shows the dependence of the pair production rate on the pump power, which has a linear trend with some saturation at higher pump powers. The saturation arises predominantly from two technical effects, detector and counting saturation and reduced performance at high powers of generic single-mode fibres for spatial filtering, both of which can be addressed in future experiments. We can now use the results from Fig. 4c to estimate experimentally the nonlinear interaction strength, Γt , that appears in equation (3). The linear fit (for the points up to a pump power of 20 mW) corresponds to a pair detection rate of 1.45 ± 0.02 pairs per second per milliwatt of 532-nm pump power (error (s.d.) calculated using Monte Carlo simulation).

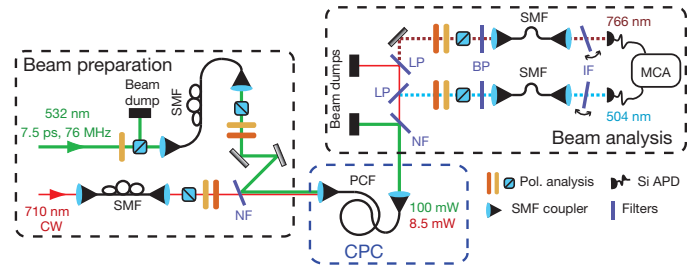


Figure 3 | Experimental set-up. There are three main stages: beam preparation (polarization state and spatial mode of input beams), CPC nonlinear interaction and beam analysis. The nonlinear medium is a standard commercial, polarization-maintaining PCF (core diameter, 1.8 μ m; nonlinearity, ~ 95 W^{-1} km^{-1}). The $\chi^{(3)}$ nonlinearity is pumped by a frequency-doubled, pulsed neodymium vanadate (Nd:YVO₄) laser (532 nm, 7.5 ps, 76 MHz) to create the desired tunable effective $\chi^{(2)}$ nonlinearity. A continuous-wave (CW), external-cavity diode laser (710 nm, $\sim 2 \times 10^5$ photons per Nd:YVO₄ pulse) provides the input state in mode a that we use to characterize the strength of the CPC interaction. From the estimated dispersion of the PCF, birefringent phase matching is satisfied for the following four-mode interaction: 532 nm (H) + 710 nm (H) \rightarrow 504 nm (V) + 766 nm (V), where H and V denote horizontal and vertical polarizations, respectively. The 532-nm and 710-nm input beams are spatially filtered with single-mode fibres (SMF) before being combined on a notch filter (NF) and coupled into the PCF. The beams emerging from the output are then spectrally separated using a range of filters (NF; LP, long-pass filter; BP, band-pass filter) and a simple monochromator (a rotating interference filter (IF)), passed through polarization analysers and, finally, detected using avalanche photodiodes (APDs) and analysed in coincidence using time-to-amplitude conversion and a multichannel analyser (MCA).

Accounting for measured losses due to coupling and optical elements in the beam analysis circuit (transmissions: $\sim 2.6\%$ in arm 1, $\sim 14.6\%$ in arm 2), this corresponds to a nonlinear interaction parameter inside the PCF of $\Gamma t \approx 5 \times 10^{-6}$ per (milliwatt)^{1/2} (estimated directly from equation (3)). For 20 mW of pump power at 532 nm, this gives an effective brightness of $\sim 10^6$ created pairs per milliwatt of effective input power at 710 nm (4.8 μ W).

We now discuss possible approaches to improving the effective $\chi^{(2)}$ nonlinearity (see Supplementary Information for more detail). For a reasonable fibre-coupled pump power of 1 W, the measured rates predict that $\Gamma t \approx 10^{-4}$, which, with a single photon in mode a , would already give pair probabilities (per photon) comparable to present state-of-the-art, high-brightness spontaneous pair sources^{3,7}. These interaction strengths could be further improved by using a lower repetition rate, so that the pulse energy can be increased without also increasing the average pump power; for example, using a system with a kilohertz repetition rate could perhaps improve Γ by another factor of up to 10^2 . Nevertheless, although we might reasonably expect future experiments to bring further technical improvements, for example by specifically engineering the nonlinear and dispersion properties of longer PCFs and matching them with the optical wavelengths, these initial results suggest that reaching the deterministic regime ($\Gamma t \approx \pi/2$) using silica might be challenging. Other materials, however, provide access to far stronger $\chi^{(3)}$ nonlinearities than does silica. For example, chalcogenide glasses, with a material $\chi^{(3)}$ up to 10^3 times that of silica, have been used to make microstructured fibres, waveguides and nanowires, which have in turn been used in ultrabroadband telecommunication devices (see ref. 30 and references therein). Combined with high refractive indices, which allow for extremely strong mode confinement in chalcogenide integrated devices, such an increase in nonlinearity should provide a promising path towards interactions strong enough to make deterministic CPC possible.

We believe that CPC is a useful technique for implementing coherent, deterministic multiphoton dynamics both for applications in quantum-enhanced technologies and for fundamental tests involving entanglement and large-scale quantum systems. CPC also provides benefits in

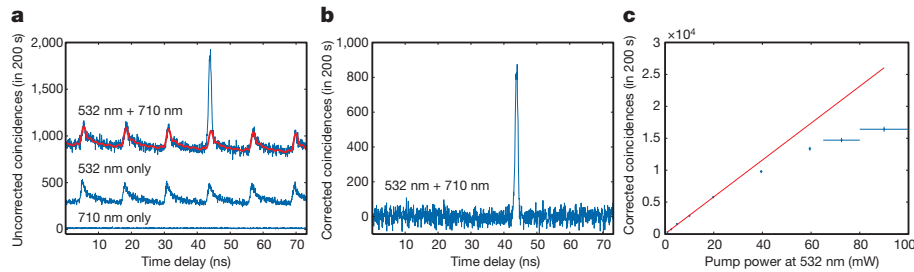


Figure 4 | Experimental results. **a**, Photon-doubling signal resulting from the four-mode nonlinear interaction that underlies our four-mode CPC. The signal (output polarization analysers both vertical) is only observed when both input beams are present (and horizontally polarized). The strong periodic background (fitted background signal shown in red; see Supplementary Information for more details) is caused by accidental coincidences from single photons created by Raman scattering from both beams, although very few such coincidences arise from just the 710-nm input because it creates very few Raman photons at 504 nm. **b**, The full MCA trace allows us to correct very precisely for this periodic background and isolate a background-subtracted signal that arises from only the CPC interaction. **c**, The pair production rate depends linearly on the pump power with some saturation at higher pump powers. The saturation arises predominantly from detector and counting

systems in which interaction strengths are substantially less than those required for deterministic operation. For example, because the photon doubler does not introduce any of the higher-order terms that limit the performance of down-conversion-based photon sources²⁹, even at low efficiencies, a CPC-based multiphoton source offers the potential for higher multiphoton rates with much lower noise terms.

Finally, we emphasize that the above four-mode example of the CPC interaction demonstrates all of the building blocks required for a CPC-based approach to photonic quantum computing. These include deterministic two-qubit entangling gates based on a novel type of effective photon-photon interaction induced by Berry's phase effects, heralded multiphoton sources (potentially entangled) with almost no higher-order terms, and efficient, low-noise single-photon detection using real-world detectors. More specifically, all of these functions can be implemented using a single, reconfigurable, eight-port device (with an input and output for each mode) controlled by choosing the appropriate inputs, for example by tuning the interaction strength with the pump power.

Received 9 June; accepted 16 August 2011.

Published online 12 October 2011.

1. Clauser, J. F. & Shimony, A. Bell's theorem: experimental tests and implications. *Rep. Prog. Phys.* **41**, 1881–1927 (1978).
2. Barreiro, J. T., Langford, N. K., Peters, N. A. & Kwiat, P. G. Generation of hyperentangled photon pairs. *Phys. Rev. Lett.* **95**, 260501 (2005).
3. Fedrizzi, A., Herbst, T., Poppe, A., Jennewein, T. & Zeilinger, A. A wavelength-tunable fiber-coupled source of narrowband entangled photons. *Opt. Express* **15**, 15377–15386 (2007).
4. Cohen, O. *et al.* Tailored photon-pair generation in optical fibers. *Phys. Rev. Lett.* **102**, 123603 (2009).
5. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
6. DiVincenzo, D. P. & Loss, D. Quantum information is physical. *Superlattices Microstruct.* **23**, 419–432 (1998).
7. Fulconis, J., Alibart, O., Wadsworth, W. J., Russell, P. S. & Rarity, J. G. High brightness single mode source of correlated photon pairs using a photonic crystal fiber. *Opt. Lett.* **13**, 7572–7582 (2005).
8. Holmes, C. A. & Milburn, G. J. Parametric self pulsing in a quantum optomechanical system. *Fortschr. Phys.* **57**, 1052–1063 (2009).
9. Chang, D. E., Safavi-Naeini, A. H., Hafezi, M. & Painter, O. Slowing and stopping light using an optomechanical crystal array. *N. J. Phys.* **13**, 023003 (2011).
10. Thompson, J. D. *et al.* Strong dispersive coupling of a high-finesse cavity to a micromechanical membrane. *Nature* **452**, 72–75 (2008).
11. Moon, K. & Girvin, S. M. Theory of microwave parametric down-conversion and squeezing using circuit QED. *Phys. Rev. Lett.* **95**, 140504 (2005).
12. Marquardt, F. Efficient on-chip source of microwave photon pairs in superconducting circuit QED. *Phys. Rev. B* **76**, 205416 (2007).
13. Kok, P. *et al.* Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.* **79**, 135–174 (2007).
14. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).

saturation and the reduced performance at high powers of the generic SMFs used to spatially filter the 532-nm beam before it is coupled into the PCF. In particular, the detector saturation results mainly from unwanted Raman scattering, and it should be possible to suppress this effect drastically by cooling the PCF. Each point corresponds to a single 200-s integration. Vertical error bars represent statistical errors (s.d.) given Poisson-distributed coincidence counts. Horizontal error bars represent the worst case uncertainty (range) in pump power inside the PCF resulting from drift in fibre coupling during the course of the measurement. The linear fit (for the points up to 20 mW of pump power) corresponds to a pair detection rate (for 4.8 μ W of effective power in the 710-nm input) of 1.45 ± 0.02 pairs per second per milliwatt of 532-nm pump power (the uncertainty (s.d.) was determined by Monte Carlo simulation with Poisson-distributed noise).

15. Walther, P. *et al.* Experimental one-way quantum computing. *Nature* **434**, 169–176 (2005).
16. Prevedel, R. *et al.* High-speed linear optics quantum computing using active feed-forward. *Nature* **445**, 65–69 (2007).
17. Munro, W. J., Nemoto, K., Beausoleil, R. G. & Spiller, T. P. High-efficiency quantum-nondemolition single-photon-number-resolving detector. *Phys. Rev. A* **71**, 033819 (2005).
18. Nemoto, K. & Munro, W. J. Nearly deterministic linear optical controlled-NOT gate. *Phys. Rev. Lett.* **93**, 250502 (2004).
19. Franson, J. D., Jacobs, B. C. & Pittman, T. B. Quantum computing using single photons and the Zeno effect. *Phys. Rev. A* **70**, 062302 (2004).
20. Shapiro, J. H. & Razavi, M. Continuous-time cross-phase modulation and quantum computation. *N. J. Phys.* **9**, 16 (2007).
21. Leung, P. M., Munro, W. J., Nemoto, K. & Ralph, T. C. Spectral effects of strong $\chi^{(2)}$ nonlinearity for quantum processing. *Phys. Rev. A* **79**, 042307 (2009).
22. Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).
23. Pancharatnam, S. Generalized theory of interference, and its applications—i. *Proc. Indian Acad. Sci. A* **44**, 247–262 (1956).
24. Rauch, H. *et al.* Verification of coherent spinor rotation of fermions. *Phys. Lett. A* **54**, 425–427 (1975).
25. VanDevender, A. P. & Kwiat, P. G. High-speed transparent switch via frequency upconversion. *Opt. Express* **15**, 4677–4683 (2007).
26. Koshino, K. Down-conversion of a single photon with unit efficiency. *Phys. Rev. A* **79**, 013804 (2009).
27. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493–R2496 (1995).
28. Zeilinger, A., Horne, M. A. & Greenberger, D. M. in *Proc. Squeezed States Quantum Uncertainty* (eds Han, D., Kim, Y. S. & Zachary, W. W.) 73–81 (NASA Conference Publication 3135, NASA, 1992).
29. Barbieri, M. *et al.* Parametric downconversion and optical quantum gates: two company, four a crowd. *J. Mod. Opt.* **56**, 209–214 (2009).
30. Eggleston, B. J., Luther-Davies, B. & Richardson, K. Chalcogenide photonics. *Nature Photon.* **5**, 141–148 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to acknowledge discussions with T. Jennewein, A. Fedrizzi, D. R. Austin, T. Paterek, B. J. Smith, W. J. Wadsworth, M. Halder, J. G. Rarity, F. Verstraete and A. G. White. This work was supported by the ERC (Advanced Grant QIT4QAD), the Austrian Science Fund (grant F4007 and an Erwin Schrödinger Fellowship), the EC (QU-ESSENCE and QAP), the Vienna Doctoral Program on Complex Quantum Systems, the John Templeton Foundation and in part by the Japanese FIRST programme and the Ontario Ministry of Research and Innovation.

Author Contributions N.K.L. and S.R. conceived the original theory and developed it with A.Z., G.J.M. and W.J.M. N.K.L., S.R., R.P. and A.Z. designed the experiment and N.K.L., S.R. and R.P. performed the experiment and carried out the data analysis. All authors contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.K.L. (nathan.langford@univie.ac.at) or A.Z. (anton.zeilinger@univie.ac.at).

PDGF signalling controls age-dependent proliferation in pancreatic β -cells

Hainan Chen¹, Xueying Gu¹, Yinghua Liu¹, Jing Wang¹, Stacey E. Wirt², Rita Bottino³, Hubert Schorle⁴, Julien Sage² & Seung K. Kim^{1,5,6}

Determining the signalling pathways that direct tissue expansion is a principal goal of regenerative biology. Vigorous pancreatic β -cell replication in juvenile mice and humans declines with age, and elucidating the basis for this decay may reveal strategies for inducing β -cell expansion, a long-sought goal for diabetes therapy. Here we show that platelet-derived growth factor receptor (Pdgfr) signalling controls age-dependent β -cell proliferation in mouse and human pancreatic islets. With age, declining β -cell Pdgfr levels were accompanied by reductions in β -cell enhancer of zeste homologue 2 (*Ezh2*) levels and β -cell replication. Conditional inactivation of the *Pdgfra* gene in β -cells accelerated these changes, preventing mouse neonatal β -cell expansion and adult β -cell regeneration. Targeted human PDGFR- α activation in mouse β -cells stimulated Erk1/2 phosphorylation, leading to *Ezh2*-dependent expansion of adult β -cells. Adult human islets lack PDGF signalling competence, but exposure of juvenile human islets to PDGF-AA stimulated β -cell proliferation. The discovery of a conserved pathway controlling age-dependent β -cell proliferation indicates new strategies for β -cell expansion.

Acquired tissue insufficiency underlies the pathogenesis of diverse human diseases; for example, absolute or relative deficits in pancreatic β -cells underlie diabetes mellitus. To expand insulin-secreting β -cells, investigators have modulated intracellular factors, including β -cell cycle regulators, or activated intrinsic regulatory pathways with extrinsic mitogenic factors^{1,2}. In mice or other experimental animals, β -cell expansion from such efforts has been transient or limited, and often complicated by uncontrolled β -cell growth, loss of defining β -cell features and impaired metabolic control^{3,4}. Likewise, induction of human islet β -cell proliferation remains an unattained goal, with studies reporting inconsistent mitogenic responses^{5,6}. Thus, we sought to identify native signalling pathways governing physiological β -cell proliferation and regeneration.

During physiological growth in humans, mice and other species, juvenile β -cells expand by self-renewal^{7,8} while maintaining their hallmark functions. Pancreatic β -cell proliferation decreases rapidly in juvenile mice and humans, then declines more slowly in adults^{9,10}. Studies indicate that age-dependent increases of p16^{INK4a}, a cyclin-dependent kinase inhibitor encoded by *Cdkn2a*, restrict proliferation of mouse and human β -cells and other tissues with ageing^{11,12}. We showed that the polycomb group protein *Ezh2* represses *Cdkn2a* and promotes β -cell proliferation in juvenile mouse islets¹³. *Ezh2* is a histone methyltransferase crucial for trimethylation of histone H3 on Lys 27 (H3K27me3), a modification mediating transcriptional repression^{14,15}. *Ezh2* expression in mouse and human β -cells declines rapidly in juveniles, then slowly in adults¹³, similar to the tempo of age-dependent decline in β -cell proliferation^{9,10}. Conditional *Ezh2* inactivation in β -cells accelerated loss of H3K27me3 repression at *Cdkn2a*, resulting in premature p16^{INK4a} expression, impaired β -cell proliferation, reduced β -cell mass and inadequate β -cell regenerative recovery after β -cell ablation¹³. We proposed that conserved regulatory pathways governing *Ezh2* expression in ageing islets could be targeted to stimulate adult β -cell proliferation.

Pdgfr signalling promotes proliferation, survival and migration in diverse cell types¹⁶. Pdgfr signalling activation stimulates DNA synthesis in cultured islets^{17,18}, but the roles of Pdgfr signalling in β -cell proliferation and expansion were unknown. Pdgfr-induced actin polymerization in fibroblasts requires *Ezh2* (ref. 19), suggesting that Pdgfr signalling might regulate *Ezh2* expression and age-dependent β -cell proliferation in islets. Here we demonstrate that a pathway regulated by Pdgfr- α , governs the expression and activity of intrinsic β -cell factors—including the retinoblastoma protein (Rb), E2f transcription factors and *Ezh2*—to control β -cell proliferation. Modulation of β -cell Pdgfr signalling attenuates age-dependent proliferative failure and *Ezh2* loss, permitting *in vivo* expansion of β -cells with regulated function. Cardinal features of this pathway are conserved in human islets, permitting conditional induction of human β -cell replication.

Age-dependent β -cell Pdgfr- α loss

To investigate whether Pdgfr signalling might regulate islet β -cell proliferation and *Ezh2* expression, we assessed Pdgfr receptor and ligand expression in islets and β -cells. Immunohistology revealed that Pdgfr- α and Pdgfr- β levels were markedly reduced in islet β -cells at 6 weeks and 6 months of age, compared to neonatal islets (Fig. 1a and Supplementary Fig. 1a). Real-time polymerase chain reaction with reverse transcription (RT-PCR) studies of β -cells purified by fluorescence-activated cell sorting (FACS) confirmed that *Pdgfra*, *Pdgfrb* and *Ezh2* messenger RNA levels were reduced in adult β -cells at 5 months compared to 2-week-old β -cells (Fig. 1b). Further RT-PCR analyses of isolated islets from wild-type mice ranging from 2 weeks to 13 months revealed an age-dependent decline of islet mRNAs encoding *Pdgfra*, *Pdgfrb* and Pdgfr ligands (Fig. 1c and Supplementary Fig. 1b–e). In contrast, islet mRNAs encoding the insulin receptor, prolactin receptor, glucocorticoid receptor and intrinsic regulators of β -cell proliferation like *Pdx1*, *Foxo1*, *Foxo3*, *Rb1* (hereafter referred to as *Rb*), *p130* (also known as *Rbl2*) and *E2f1*, did not decline in mice with advancing

¹Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Departments of Pediatrics and Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ³Department of Pediatrics, Division of Immunogenetics, Children's Hospital of Pittsburgh, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania 15213, USA. ⁴Department of Developmental Pathology, Institute of Pathology, University of Bonn Medical School, 53127 Bonn, Germany. ⁵Department of Medicine (Oncology), Stanford University School of Medicine, Stanford, California 94305, USA. ⁶Howard Hughes Medical Institute, Department of Developmental Biology, B300 Beckman Center, 279 Campus Drive, Stanford, California 94305, USA.

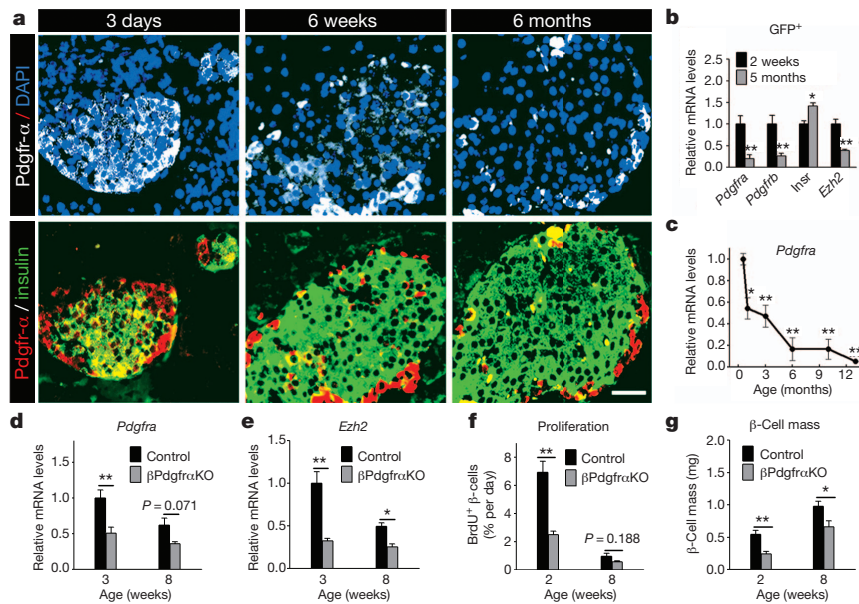


Figure 1 | Age-dependent attenuation of Pdgfr- α limits β -cell *Ezh2* expression and proliferation in neonatal and juvenile mice. **a**, Pdgfr- α and insulin immunostaining on wild-type mouse pancreas sections at indicated ages. DAPI, 4',6-diamidino-2-phenylindole. Scale bar, 25 μ m. **b**, **c**, mRNA levels of indicated genes in FACS-purified β -cells (**b**) or wild-type islets (**c**) at specified ages. $n = 3$ –5 experiments per group or time point. * $P < 0.05$,

** $P < 0.01$ compared to samples from 2 weeks (**b**) or 15 days (**c**). *Insr*, refers to mRNA encoding insulin receptor. **d**–**g**, Relative mRNA levels of *Pdgfra* (**d**) and *Ezh2* (**e**) in isolated islets, and percentage β -cell BrdU incorporation (**f**) and pancreatic β -cell mass (**g**) of β Pdgfr α KO and control mice at specified ages. $n = 3$ –5 independent islet preparations or mice per group. * $P < 0.05$, ** $P < 0.01$ for the comparison as indicated. Error bars denote s.e.m.

age (Supplementary Fig. 2). Thus, reduced β -cell levels of Pdgfr receptors and ligands corresponded with *Ezh2* loss and age-dependent reduction of β -cell replication¹³.

To assess whether premature Pdgfr loss might compromise *Ezh2*-dependent neonatal β -cell proliferation, we generated *RIP-Cre; Pdgfra^{fl/fl}* mice (abbreviated to β Pdgfr α KO) that permit conditional *Pdgfra* inactivation in islet β -cells and loss of detectable *Pdgfra* mRNA and protein expression (Fig. 1d and Supplementary Fig. 3a). Compared to controls, β -cell *Ezh2* protein and islet *Ezh2* mRNA levels were reduced whereas *p16^{INK4a}* mRNA levels were increased in β Pdgfr α KO islets (Fig. 1e and Supplementary Fig. 3a, b); in contrast, *Pdgfrb* and *Ezh1* mRNA levels were unchanged in β Pdgfr α KO islets (Supplementary Fig. 3c, d). In 2–3-week-old β Pdgfr α KO mice, we detected a threefold reduction in the β -cell proliferation rate and a 50% reduction in β -cell mass (Fig. 1f, g). β Pdgfr α KO mice at 3 weeks had mild hyperglycaemia during *ad libitum* feeding and impaired glucose tolerance in a glucose challenge test (Supplementary Fig. 3e, f). Thus, premature loss of *Pdgfra* impairs β -cell proliferation and expansion and glucose control in young mice.

Pdgfr- α controls adult β -cell regeneration

Destruction of β -cells by streptozotocin (STZ) in mice provokes β -cell proliferation with regeneration of β -cell mass^{11,13}. Following STZ exposure, *Pdgfra* and *Pdgfrb* were induced in wild-type islets (Fig. 2a, b), suggesting possible roles for Pdgfr signalling in adult β -cell regeneration. Consistent with this view, β -cell mass in control mice gradually recovered after STZ treatment, accompanied by increased β -cell *Ezh2* expression and proliferation; in contrast, β Pdgfr α KO β -cells failed to increase *Ezh2* or proliferation to restore β -cell mass (Fig. 2c–e). β Pdgfr α KO mice developed severe diabetes that persisted throughout the course of our studies (Fig. 2f). In contrast, control mice became moderately diabetic immediately after STZ challenge, but achieved restored glucose control 3–4 weeks thereafter (Fig. 2f), coinciding with increases of β -cell mass (Fig. 2e). Thus, Pdgfr- α is also required in adult β -cells for compensatory proliferation in experimental diabetes.

PDGFR- α promotes functional β -cell expansion

To test whether enhanced Pdgfr signalling might promote β -cell expansion, we exposed islets from juvenile and adult mice to physiological concentrations of platelet-derived growth factor-AA (PDGF-AA). Exposure of 3-week-old islets to PDGF-AA increased phospho-Pdgfr- α levels and *Ezh2* mRNA and protein levels, but not *Ezh1* expression (Supplementary Fig. 4a–c). In contrast, juvenile islets exposed to mitogens like insulin or prolactin did not detectably alter *Ezh2* mRNA levels (Supplementary Fig. 4d, e). Unlike in 3-week-old islets, levels of *Ezh2* mRNA and protein were not significantly increased in 7- or 9-month-old adult islets exposed to PDGF-AA (Supplementary Fig. 4c, f). PDGF-AA exposure increased β -cell proliferation in juvenile islets, as assessed by BrdU incorporation, but not in adult islets (Supplementary Fig. 4g, h). Increased *Ezh2* expression and β -cell BrdU incorporation were eliminated by simultaneous treatment with the receptor tyrosine kinase inhibitors Sunitinib²⁰ or Vargatef²¹ (Supplementary Fig. 4a, c, g, h). Thus, PDGF-AA exposure was sufficient to stimulate β -cell replication in juvenile islets, but loss of competence for Pdgfr prevented this response to PDGF-AA in adult islets.

To investigate whether sustained PDGFR signalling permitted adult β -cell expansion *in vivo*, we intercrossed the *RIP-Cre* strain with mice harbouring a human *PDGFRA*^{D842V} allele encoding a ligand-independent activated receptor inserted at the *ROSA26* (*R26*) locus²². A loxP-flanked transcriptional stop sequence permits Cre-dependent *PDGFRA*^{D842V} expression and activation of PDGF signalling *in vivo*²². In contrast to littermate controls, mice with the *RIP-Cre; R26-PDGFR α ^{D842V}* genotype (abbreviated to β PDGFR α Tg) expressed human *PDGFRA* mRNA and PDGFR- α (D842V) protein in islet β -cells (Fig. 3a and Supplementary Fig. 5a). Expression of *R26-PDGFR α ^{D842V}* did not affect endogenous islet *Pdgfra* mRNA levels (Fig. 3a) or β -cell expansion and glucose control in neonatal or juvenile β PDGFR α Tg mice (Fig. 3b, c and Supplementary Figs 5b and 6a–d). By 3 months, β -cell *Ezh2* expression had declined in control mice, but β -cell *Ezh2* expression in β PDGFR α Tg mice was maintained at elevated levels, a difference that persisted for 8–14 months (Fig. 3e, f). mRNA levels of islet *p16^{INK4a}* and *p19^{Arf}*, both targets¹³ of *Ezh2*, were

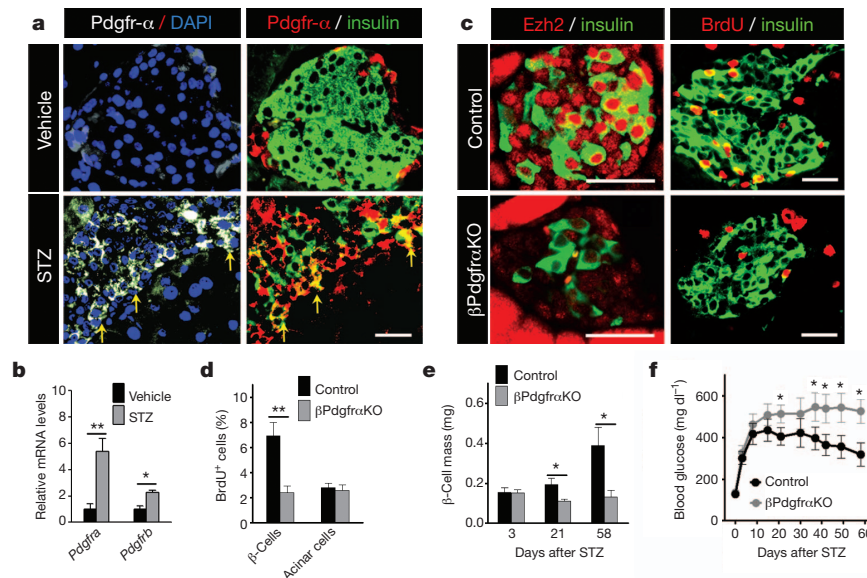


Figure 2 | *Pdgfra* loss impairs β -cell regeneration in STZ-induced diabetes. **a, b**, Immunostaining of Pdgfr- α and insulin on pancreas sections (**a**), and relative mRNA levels of *Pdgfra* and *Pdgfrb* in islets (**b**) from 8-week-old wild-type mice 1 week after vehicle or STZ (100 mg kg $^{-1}$ body weight) treatment. $n = 3$ –5 mice per group. **a**, Yellow arrows mark Pdgfr- α induction in a subset of insulin $^{+}$ cells. **c**, Representative images showing immunostaining for insulin, Ezh2 and BrdU on pancreatic sections from indicated 6–7-week-old mice 3

weeks after STZ (150 mg kg $^{-1}$ body weight) treatment. **d**, Percentage of BrdU $^{+}$ insulin $^{+}$ cells (β -cells) or BrdU $^{+}$ acinar cells 3 weeks after STZ challenge. $n = 3$ or 6 mice per group. **e, f**, Pancreatic β -cell mass (**e**) and blood glucose (**f**) levels during *ad libitum* feeding after STZ treatment. $n = 3$ –8 mice (**e**) or 18–21 mice (**f**) per time point per group. * $P < 0.05$, ** $P < 0.01$ for the comparison as indicated or versus control. Scale bars, 25 μ m. Error bars denote s.e.m.

reduced in β PDGFR α Tg mice compared to those in littermate controls (Supplementary Fig. 6f, g). In contrast, islet mRNA levels of *Ezh1* in β PDGFR α Tg and control mice were indistinguishable (Supplementary Fig. 6h). β -Cell proliferation in 14-month-old β PDGFR α Tg mice was increased ninefold compared to age-matched controls, a level of proliferation seen in 3-month-old control β -cells (Fig. 3d, f). β -Cell

mass was increased in β PDGFR α Tg mice at 3 months, and remained increased at 14 months compared to age-matched controls (Fig. 3b). In contrast, total pancreatic mass and islet architecture in β PDGFR α Tg mice and controls were indistinguishable (Supplementary Figs 5c and 6b). Thus, Pdgfr signalling activation is sufficient to sustain adult β -cell expansion *in vivo*.

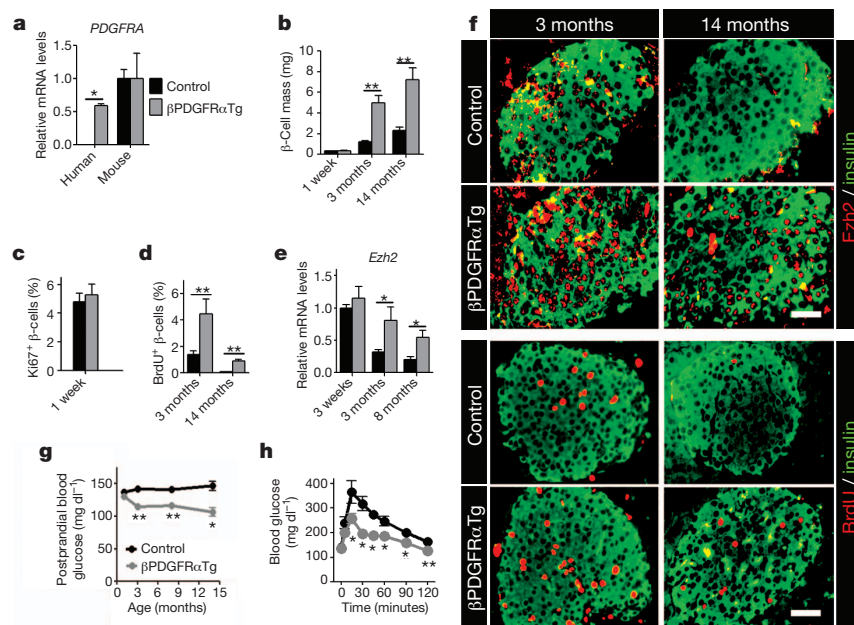


Figure 3 | Activated PDGFR- α delays age-dependent *Ezh2* loss and replication failure in pancreatic β -cells. **a**, Relative mRNA levels of human *PDGFRA* and mouse *Pdgfra* in islets from littermate control and β PDGFR α Tg mice at 3 months of age. $n = 3$ –7 mice per group. **b–d**, Pancreatic β -cell mass (**b**), and β -cell proliferation, assessed by Ki67 expression (**c**) or BrdU incorporation (**d**) in β PDGFR α Tg and control mice at indicated ages. $n = 3$ –5 mice per group. **e, f**, Relative mRNA levels of *Ezh2* (**e**) in islets, and

immunostaining (**f**) for Ezh2, BrdU and insulin on pancreatic sections from control and β PDGFR α Tg mice at indicated ages. $n = 3$ –7 mice per group. Scale bars, 25 μ m. **g**, Postprandial blood glucose levels in mice at indicated ages. $n = 14$ –24 mice per group per time point. **h**, Glucose tolerance assessed in 14-month-old control and β PDGFR α Tg mice. $n = 6$ or 10 mice per group. * $P < 0.05$, ** $P < 0.01$ in comparisons indicated. Error bars denote s.e.m.

Prior studies indicate that β -cell expansion stimulated by mitogens can be complicated by disrupted β -cell function and impaired metabolic control^{3,4}. To assess the physiological impact of prolonged adult β -cell proliferation in β PDGFR α Tg mice, we measured systemic blood glucose and insulin levels. Blood glucose levels during fasting and *ad libitum* feeding and responses to insulin challenge were indistinguishable in β PDGFR α Tg mice and littermate controls (Supplementary Fig. 6c–e). During re-feeding after overnight fast, β PDGFR α Tg mice had a 25% reduction of blood glucose (Fig. 3g), associated with a modest elevation of plasma insulin levels (Supplementary Fig. 6i), and both changes were sustained for up to 14 months. After overnight fast, glucose clearance was enhanced in glucose-challenged 3- or 14-month-old β PDGFR α Tg mice (Fig. 3h and Supplementary Fig. 6j). Thus, PDGFR α Tg mice maintained normal glucose-regulated insulin release and glycaemic control, suggesting that β -cell function remained regulated in these mice.

PDGF-induced β -cell expansion requires *Ezh2*

To test whether PDGFR- α -induced β -cell proliferation and expansion required *Ezh2*, we generated mice permitting conditional inactivation¹³ of *Ezh2* in β PDGFR α Tg β -cells. Intercrosses generated *RIP-Cre; Ezh2^{fl/f}; R26-PDGFR α ^{D842V}* mice (abbreviated to β Ezh2KO-R α Tg) and littermate or sibling *RIP-Cre; Ezh2^{fl/+}; R26-PDGFR α ^{D842V}* mice (abbreviated to β Ezh2HET-R α Tg) and *RIP-Cre; Ezh2^{fl/f}* mice (abbreviated to β Ezh2KO). Islet *PDGFR α ^{D842V}* mRNA levels in both β Ezh2KO-R α Tg and β Ezh2HET-R α Tg mice were elevated and indistinguishable (Fig. 4a). By contrast, increased islet *Ezh2* mRNA levels and β -cell *Ezh2* protein were not observed in β Ezh2KO-R α Tg or β Ezh2KO littermates (Fig. 4b and Supplementary Fig. 7), confirming Cre-mediated deletion of *Ezh2* in these mice. As expected, β -cell BrdU incorporation and mass were increased in β Ezh2HET-R α Tg mice (Fig. 4c, d and Supplementary Fig. 7) compared to controls (either *Ezh2^{fl/f}*, *R26-PDGFR α ^{D842V}* or *Ezh2^{fl/+}; R26-PDGFR α ^{D842V}* genotype). However, neither β -cell BrdU incorporation nor mass were increased in β Ezh2KO-R α Tg mice compared to controls (Fig. 4c, d and Supplementary Fig. 7). Consistent with these findings, we detected modest postprandial blood glucose reduction and increased blood insulin

levels in β Ezh2HET-R α Tg mice, but not in β Ezh2KO-R α Tg mice (Fig. 4e, f). Thus, *Ezh2* is required for β -cell expansion and metabolic changes from PDGFR- α activation in β -cells.

β -Cell PDGFR controls *Ezh2* via Erk and Rb/E2f

To elucidate the signalling basis of altered β -cell proliferation in β PDGFR α Tg, β Pdgfr α KO and ageing wild-type mice, we investigated Pdgf signal transduction factors. Pdgf receptors activate signalling elements¹⁶, including the mitogen-activated protein kinase/extracellular signal-regulated kinase (Mapk/Erk), phosphatidylinositol 3-kinase (PI3K)/Akt and phospholipase PLC- γ . *Ezh2* induction in juvenile wild-type islets exposed to PDGF-AA was blocked by U0126, which inhibits Erk1/2 activation²³, but not by LY294002 or U-73122, inhibitors of PI3K or PLC- γ signalling (Supplementary Fig. 8a). Similarly, increased β -cell BrdU incorporation in islets exposed to PDGF-AA was blocked by U0126, but not by LY294002 (Supplementary Fig. 8b), indicating that Erk mediates β -cell Pdgf signalling responses. We also detected increased phosphorylation of Erk1/2, but not phosphorylation of Akt or PLC- γ , in β PDGFR α Tg islets compared to control islets (Fig. 5a, b). Immunohistology confirmed increased phospho-Erk1/2 levels in β -cells of β PDGFR α Tg islets compared to controls (Fig. 5c). PDGF/ERK signalling activates cell cycle regulators to stimulate proliferation^{24,25}, and β PDGFR α Tg islets had elevated levels of *Ccnd1* mRNA encoding cyclin D1, and increased β -cells with nuclear cyclin D1 and phosphorylated Rb protein (Fig. 5c–e and Supplementary Fig. 9a). In contrast, loss of β -cell Pdgf signalling in β Pdgfr α KO mice diminished islet levels of *Ccnd1* mRNA, and reduced numbers of phospho-Erk1/2⁺ and phospho-Rb⁺ β -cells (Supplementary Fig. 9b, c). Rb protein phosphorylation regulates association of the E2f1 transcriptional activator and E2f4 transcription repressor with their targets^{26,27}. We identified two conserved candidate E2f-binding elements near the *Ezh2* promoter (Supplementary Fig. 10), and chromatin immunoprecipitation (ChIP) studies showed increased E2f1 association and reduced E2f4 association with these *Ezh2* elements in β PDGFR α Tg islets (Fig. 5f, g). In wild-type mice, levels of phospho-Erk1/2 and phospho-Rb declined with age in islet β -cells at a tempo similar to declining *Ezh2* and *Pdgfra* expression (Supplementary Fig. 11a, b). ChIP studies also revealed that association of E2f1 with *Ezh2* decreased, whereas association of E2f4 with *Ezh2* elements increased in islets with age (Supplementary Fig. 11c, d). Thus, age-dependent attenuation of β -cell Pdgfr- α altered Erk and Rb/E2f signalling, resulting in reduced *Ezh2* expression.

To verify that Rb/E2f signalling regulates β -cell *Ezh2*, we examined triple mutant *ROSA-CreER^{T2}; Rb^{fl/f}; p130^{fl/f}; p107^{-/-}* mice (RbTriKO) lacking Rb family proteins²⁸. After exposure to tamoxifen, β -cells in adult RbTriKO mice, lacking detectable Rb, p130 and p107, had increased *Ezh2* protein and mRNA levels, compared to those from tamoxifen-exposed *Rb^{fl/f}; p130^{fl/f}; p107^{-/-}* control littermates (Supplementary Fig. 12a–c). Islet ChIP studies revealed increased E2f1 association, and reduced E2f4 association with *Ezh2* cis-regulatory elements (Supplementary Fig. 12d, e). Consistent with these findings, BrdU incorporation in RbTriKO β -cells was increased by 15-fold, culminating in a 400% increase of β -cell mass and hypoglycaemia that progressively worsened until death (Supplementary Fig. 12f–h). Thus, Rb proteins constrain *Ezh2* expression and adult islet cell growth.

PDGF control of human β -cell proliferation

To investigate the relevance of our studies to human β -cells, we assessed PDGF signalling and competence in human islets. Like in juvenile mice, PDGFR- α protein and phospho-ERK1/2 levels were readily detected in islets from young human donors (Fig. 6a, b). We also detected abundant phospho-RB-Ser 780 and EZH2 protein in β -cell nuclei from young islets (Fig. 6b). In contrast, PDGFR- α was only detected in non- β -cells of islets from adult human donors, with little to no detectable phospho-ERK1/2, phospho-RB or EZH2 in

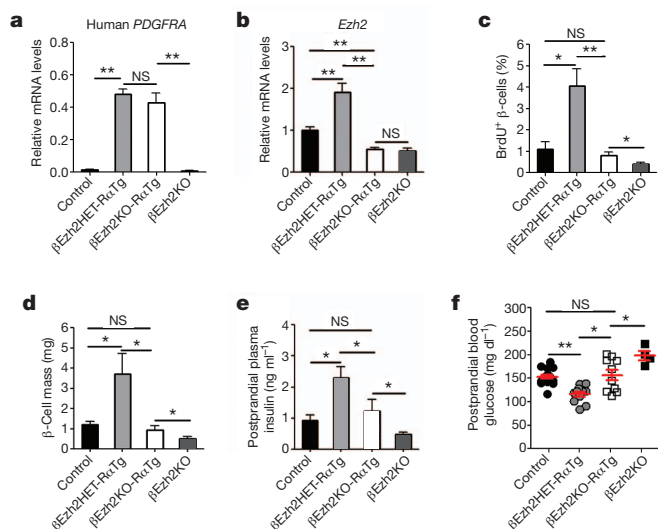


Figure 4 | PDGFR- α promotes β -cell expansion through *Ezh2*. **a, b**, mRNA levels for human *PDGFRA* (**a**) and mouse *Ezh2* (**b**) in islets from mice with indicated genotypes at 3–4 months of age. $n = 3$ –5 mice per group. **c–f**, β -Cell BrdU incorporation (**c**), β -cell mass (**d**), postprandial plasma insulin (**e**) and blood glucose (**f**) levels in 3–4-month-old mice with indicated genotypes. Each dot in **f** represents a measurement from an individual mouse. $n = 4$ –5 mice for **c** and **d**, and $n = 5$ –18 mice for **e** and **f** per genotype. * $P < 0.05$, ** $P < 0.01$. NS, not significant in comparisons indicated. Error bars denote s.e.m.

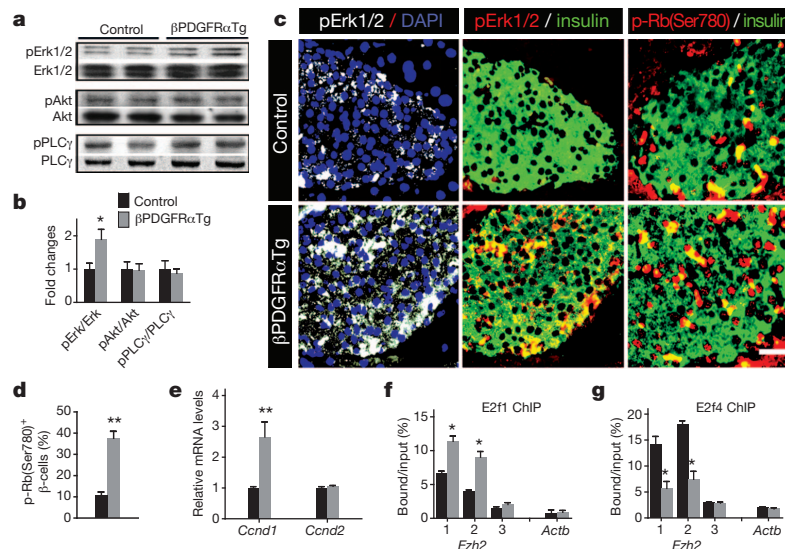


Figure 5 | Pdgfr signalling governs Erk and Rb/E2f regulation of *Ezh2* in islet β-cells. **a**, Western blot analysis of total and phosphorylated Erk1/2, Akt and PLCγ in islet proteins from 3–4-month-old βPDGFRαTg and littermate control mice. Similar results were obtained from multiple independent experiments. **b**, Relative phosphorylated protein level compared to total protein was quantified by densitometry. **c**, Immunostaining for phospho-Erk1/2 (pErk1/2), phospho-Rb-Ser 780 (p-Rb(Ser780)) and insulin in 3-month-old littermate control and βPDGFRαTg pancreatic sections. Scale bar, 25 μm. **d**, Percentage of phospho-Rb-Ser 780⁺ insulin⁺ cells from the indicated mice

quantified by morphometry. *n* = 4 mice per group. **e**, Cyclin D1 (*Ccnd1*) or cyclin D2 (*Ccnd2*) mRNA levels in 3-month-old control and βPDGFRαTg islets. *n* = 4 or 6 mice per group. **f**, **g**, ChIP analyses of the *Ezh2* and β-actin (*Actb*) loci with antibodies to E2f1 (**f**) and E2f4 (**g**) using the indicated amplicons (see Supplementary Information) in 3–4-month-old littermate control and βPDGFRαTg islets. *n* = 3–4 independent experiments per antibody with independent islet samples. **P* < 0.05, ***P* < 0.01 for control versus βPDGFRαTg. Error bars denote s.e.m.

β-cells (Fig. 6a, b), suggesting that PDGF signalling attenuation is a conserved feature of ageing islet β-cells. mRNA levels of *EZH2*, but not *EZH1*, were increased in juvenile human islets exposed to PDGF-AA, an effect blocked by Sunitinib or U0126 (Fig. 6c and Supplementary Fig. 13a). In contrast, basal *EZH2* mRNA levels were lower in adult islets and not induced by PDGF-AA (Fig. 6c). Similar to juvenile mouse islets with activated Pdgfr signalling, ChIP studies demonstrated PDGF-AA-stimulated association of E2F1 with the human *EZH2* locus, an effect blocked by Sunitinib (Fig. 6d). Thus, evolutionarily conserved elements of the PDGF signalling pathway govern *EZH2* expression and β-cell cycle regulators in human islets. Proliferating BrdU⁺ cells in juvenile islets stimulated by PDGF-AA were predominantly β-cells, verified by immuno-localization of BrdU⁺ nuclei within insulin⁺ cells and PDX1⁺ nuclei with confocal microscopy (Fig. 6f and Supplementary Fig. 13b, c). Compared to vehicle-exposed controls, juvenile islets exposed to PDGF-AA had a sixfold increase of β-cell BrdU incorporation, an effect eliminated by simultaneous exposure to Sunitinib or U0126 (Fig. 6e, f). In contrast, PDGF-AA did not alter β-cell proliferation in adult human islets (Fig. 6e, f). Thus, dynamic PDGF signalling competence may regulate declining β-cell proliferation in ageing human islets.

Discussion

The elucidation and control of mechanisms governing pancreatic β-cell proliferation could transform treatments for diseases like diabetes. However, prior attempts to expand islets by modulating extrinsic or intrinsic growth regulators have been bedevilled by limited proliferation with loss of defining β-cell features^{3,4}, indicating disruption of mechanisms that preserve fate in proliferating β-cells. The maintenance of cell fate and function in proliferating cells is the essence of epigenetic regulation. Here we focused on identifying the principal elements of a native signalling pathway regulating *Ezh2*, an essential epigenetic regulator of β-cell proliferation¹³. We found that β-cell Pdgfr signalling was required for *Ezh2* induction, to sustain both physiological and regenerative β-cell proliferation.

During islet β-cell regeneration (this study), and in maternal islets in pregnancy (H.C. and S.K.K., unpublished results), we observed a

twofold increase of islet *Ezh2* mRNA levels. However, pathological loss of all Rb family proteins in RbTriKO islets led to a fourfold increase in *Ezh2* mRNA, and a 15-fold increase in β-cell proliferation. RbTriKO mice became hypoglycaemic and developed other islet and metabolic phenotypes reminiscent of insulinoma pathogenesis. Thus, the physiological expansion of β-cells maintaining regulated function may require mechanisms limiting *Ezh2* induction. Compared with β-cells from βEzh2KO mice, we also observed a modest enhancement of proliferation and expansion after PDGFR-α activation in β-cells lacking *Ezh2*. Insulin or prolactin, two mitogens capable of promoting β-cell proliferation^{1,2,29}, did not induce β-cell *Ezh2* expression in cultured islets. Thus, our findings here support the view that *Ezh2*-independent mechanisms regulate β-cell proliferation. Likewise, decline of β-cell proliferation in ageing βPDGFRαTg mice suggests that PDGF-independent pathways restrict β-cell expansion.

Proliferative responses to PDGF decline in some ageing cells³⁰, and progenitor cell self-renewal is impaired by Pdgfr signalling disruption³¹. However, the molecular basis for these observations and the roles of Pdgfr signalling in islet β-cell growth and development were not established^{17,18,32}. Here we showed that loss of β-cell Pdgfr-α signalling accelerated β-cell replication failure both in juvenile mice and in adult mice after conditional β-cell ablation. *In vivo* activation of β-cell PDGFR-α signalling increased *Ezh2* expression and mitigated age-dependent decline of β-cell proliferation. Crucially, conditional *Ezh2* inactivation in β-cells with activated PDGFR-α signalling prevented these changes. Thus, our conditional loss- and gain-of-function studies reveal how Pdgfr signalling regulates age-dependent physiological β-cell proliferation and expansion. Our work also suggests that Mapk/Erk regulation links Pdgfr activation to *Ezh2* induction and β-cell proliferation, consistent with studies demonstrating Erk regulation of *Ezh2* in cancer cells³³, and Erk activation that accompanies mouse β-cell expansion in pregnancy³⁴. Further studies should clarify how Erk and other factors might integrate β-cell responses to Pdgfr or other mitogens³⁵.

Although the mechanisms underlying pancreatic β-cell expansion in physiological settings may differ between species^{36,37}, age-dependent restriction of β-cell proliferation is a conserved feature in mice and

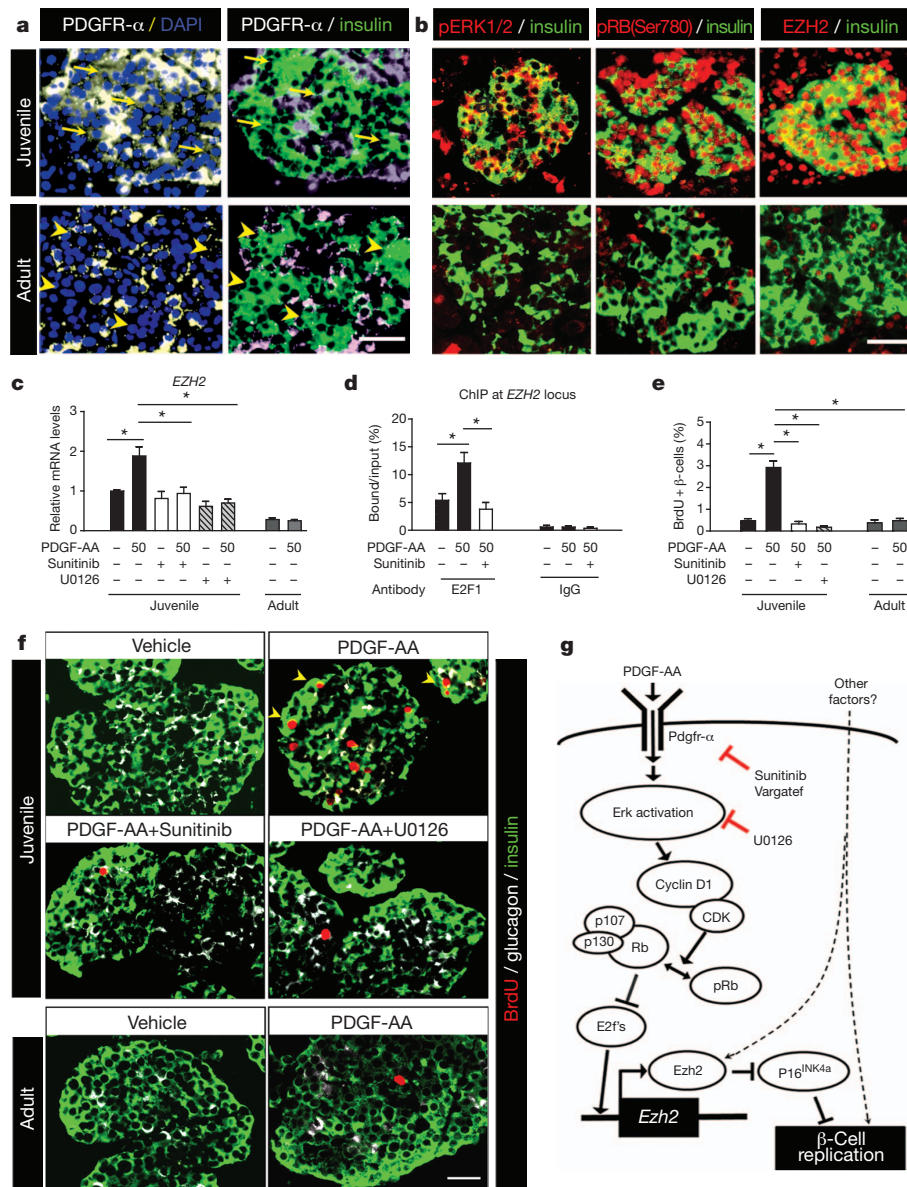


Figure 6 | PDGFR- α regulates human β -cell *EZH2* expression and proliferation. **a, b**, Representative images showing immunostaining for PDGFR- α (**a**), phospho-ERK1/2 (pERK1/2), phospho-RB-Ser 780 (pRB(Ser780)), *EZH2* (**b**) and insulin (**a** and **b**) on pancreatic sections from juvenile and adult human subjects. PDGFR- α was detected in juvenile β -cells (arrows) but not in adult β -cells (arrowheads). Scale bars, 25 μ m. **c–f**, Assessment of effects on human juvenile or adult islets after exposure to PDGF-AA (50 ng ml⁻¹) for 2 days, with or without Sunitinib (2 μ M) or U0126 (10 μ M) co-treatment. **c**, Islet *EZH2* mRNA levels after the indicated treatments.

humans^{9,10}. Here, we provide unprecedented evidence that a conserved signalling pathway restricts β -cell proliferation in mice and humans. We suggest that β -cell competence for signalling factors like PDGF can underlie dynamic islet responses to mitogens. Such age-dependent signalling competence might explain earlier findings of inconsistent human β -cell responses to other mitogens^{5,6}. PDGF ligands are measurable in serum, and regulation of such circulating signalling agonists may provide an additional level of growth control in competent islet β -cells. Prior studies correlated increased *PDGFR* and *EZH2* expression in human endocrine neoplasias³⁸, and disrupted PDGF signalling in β -cells of humans with type 2 diabetes³⁹. Thus, our demonstration that PDGF signalling regulates *EZH2*, *CDKN2A* and β -cell proliferation in human islets combines several observations into a new molecular model for physiological and pathological β -cell growth. The factors

n = 3–5 independent experiments. **d**, *EZH2* locus ChIP analysis with anti-E2F1 antibody or IgG in human juvenile islets. *n* = 3–4 for E2F1, *n* = 2 for IgG. **e, f**, Human islet β -cell proliferation changes after the indicated treatments. Average percentage of BrdU⁺ insulin⁺ cells (**e**) was quantified by morphometry from sectioned islets immunostained (**f**) for insulin (green), glucagon (white) and BrdU (red). *n* = 3–6 independent experiments. Scale bar, 25 μ m. **g**, Illustration summarizing how Pdgfr- α signalling regulates β -cell *Ezh2* and proliferation by activating Erk/Rb/E2f pathways sensitive to the indicated inhibitors. **P* < 0.05, ***P* < 0.01 as indicated. Error bars denote s.e.m.

governing β -cell expression of PDGF signalling regulators like PDGFR- α remain unknown, but we speculate that the identification of such factors may be useful for modulating human β -cell growth and function in diabetes and cancer.

METHODS SUMMARY

Animal studies and mouse islet culture experiments. Details for various mouse strain generation, *in vivo* manipulation and measurement, mouse islet isolation and culture and FACS-based β -cell purification are provided in Supplementary Data. **Histology, immunohistochemistry and molecular studies.** Standard histology and immunostaining protocols for various proteins on pancreas or islet sections were described previously¹³. β -Cell mass and proliferation were morphometrically quantified¹³ by independent researchers blinded to sample identity. Details of these studies and real-time RT-PCR, western blot and ChIP analyses are provided in Supplementary Data.

Human pancreas and islet studies. Four juvenile and four adult pancreases, and eight juvenile and three adult individual islet preparations from healthy human deceased organ donors (mean age 3.0 ± 0.6 years for juveniles; mean age 44.3 ± 5.3 years for adults) were used in this study. As in the mouse studies, similar procedures and experiments were performed on human pancreatic sections or islet cultures as described in Supplementary Data.

Statistical analyses. Results are shown as mean \pm standard error of the mean (s.e.m.). Statistical analysis was performed by one-way ANOVA, two-way repeated measures ANOVA, or unpaired Student's *t*-test, as appropriate, with significance set at $P < 0.05$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 June; accepted 24 August 2011.

Published online 12 October 2011.

- Heit, J. J., Karnik, S. K. & Kim, S. K. Intrinsic regulators of pancreatic β -cell proliferation. *Annu. Rev. Cell Dev. Biol.* **22**, 311–338 (2006).
- Vasavada, R. C. *et al.* Growth factors and β -cell replication. *Int. J. Biochem. Cell Biol.* **38**, 931–950 (2006).
- Vasavada, R. C. *et al.* Targeted expression of placental lactogen in the β -cells of transgenic mice results in β -cell proliferation, islet mass augmentation, and hypoglycemia. *J. Biol. Chem.* **275**, 15399–15406 (2000).
- Garcia-Ocana, A. *et al.* Hepatocyte growth factor overexpression in the islet of transgenic mice increases β -cell proliferation, enhances islet mass, and induces mild hypoglycemia. *J. Biol. Chem.* **275**, 1226–1232 (2000).
- Beattie, G. M. *et al.* A novel approach to increase human islet cell mass while preserving β -cell function. *Diabetes* **51**, 3435–3439 (2002).
- Parnaud, G. *et al.* Proliferation of sorted human and rat β -cells. *Diabetologia* **51**, 91–100 (2008).
- Dor, Y. *et al.* Adult pancreatic β -cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41–46 (2004).
- Kushner, J. A. *et al.* Cyclins D2 and D1 are essential for postnatal pancreatic β -cell growth. *Mol. Cell. Biol.* **25**, 3752–3762 (2005).
- Teta, M. *et al.* Very slow turnover of β -cells in aged adult mice. *Diabetes* **54**, 2557–2567 (2005).
- Meier, J. J. *et al.* β -Cell replication is the primary mechanism subserving the postnatal expansion of β -cell mass in humans. *Diabetes* **57**, 1584–1594 (2008).
- Krishnamurthy, J. *et al.* p16^{INK4a} induces an age-dependent decline in islet regenerative potential. *Nature* **443**, 453–457 (2006).
- Zindy, F. *et al.* Expression of the p16^{INK4a} tumor suppressor versus other INK4 family members during mouse development and aging. *Oncogene* **15**, 203–211 (1997).
- Chen, H. *et al.* Polycomb protein Ezh2 regulates pancreatic β -cell Ink4a/Arf expression and regeneration in diabetes mellitus. *Genes Dev.* **23**, 975–985 (2009).
- Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
- van der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nature Genet.* **23**, 474–478 (1999).
- Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225 (2000).
- Swenne, I. *et al.* Effects of platelet-derived growth factor and somatomedin-C/insulin-like growth factor I on the deoxyribonucleic acid replication of fetal rat islets of Langerhans in tissue culture. *Endocrinology* **122**, 214–218 (1988).
- Welsh, M. *et al.* Coexpression of the platelet-derived growth factor (PDGF) B chain and the PDGF beta receptor in isolated pancreatic islet cells stimulates DNA synthesis. *Proc. Natl Acad. Sci. USA* **87**, 5807–5811 (1990).
- Su, I. *et al.* Polycomb group protein Ezh2 controls actin polymerization and cell signaling. *Cell* **121**, 425–436 (2005).
- Mendel, D. B. *et al.* *In vivo* antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. *Clin. Cancer Res.* **9**, 327–337 (2003).
- Hilberg, F. *et al.* BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. *Cancer Res.* **68**, 4774–4782 (2008).
- Moenning, A. *et al.* Sustained platelet-derived growth factor receptor α signaling in osteoblasts results in craniosynostosis by overactivating the phospholipase C- γ pathway. *Mol. Cell. Biol.* **29**, 881–891 (2009).
- Duncia, J. V. *et al.* MEK inhibitors: the chemistry and biological activity of U0126, its analogs, and cyclization products. *Bioorg. Med. Chem. Lett.* **8**, 2839–2844 (1998).
- Uhrbom, L., Nerio, E. & Holland, E. C. Dissecting tumor maintenance requirements using bioluminescence imaging of cell proliferation in a mouse glioma model. *Nature Med.* **10**, 1257–1260 (2004).
- Furstoss, O., Manes, G. & Roche, S. Cyclin E and cyclin A are likely targets of Src for PDGF-induced DNA synthesis in fibroblasts. *FEBS Lett.* **526**, 82–86 (2002).
- Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**, 5323–5335 (2003).
- Moberg, K., Starz, M. A. & Lees, J. A. E2F4 switches from p130 to p107 and pRB in response to cell cycle reentry. *Mol. Cell. Biol.* **16**, 1436–1449 (1996).
- Viatour, P. *et al.* Hematopoietic stem cell quiescence is maintained by compound contributions of the retinoblastoma gene family. *Cell Stem Cell* **3**, 416–428 (2008).
- Brelje, T. C., Parsons, J. A. & Sorenson, R. L. Regulation of islet β -cell proliferation by prolactin in rat islets. *Diabetes* **43**, 263–273 (1994).
- Scherping, S. C. Jr *et al.* Effect of growth factors on the proliferation of ligament fibroblasts from skeletally mature rabbits. *Connect. Tissue Res.* **36**, 1–8 (1997).
- Pallante, B. A. *et al.* Bone marrow Oct3/4⁺ cells differentiate into cardiac myocytes via age-dependent paracrine mechanisms. *Circ. Res.* **100**, e1–e11 (2007).
- LeBras, S., Czernichow, P. & Scharfmann, R. A search for tyrosine kinase receptors expressed in the rat embryonic pancreas. *Diabetologia* **41**, 1474–1481 (1998).
- Fujii, S. *et al.* MEK–ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes. *Oncogene* doi:10.1038/onc.2011.118 (18 April 2011).
- Gupta, R. K. *et al.* Expansion of adult β -cell mass in response to increased metabolic demand is dependent on HNF-4 α . *Genes Dev.* **21**, 756–769 (2007).
- Miettinen, P. J. *et al.* Downregulation of EGF receptor signaling in pancreatic islets causes diabetes due to impaired postnatal β -cell growth. *Diabetes* **55**, 3299–3308 (2006).
- Butler, A. E. *et al.* Adaptive changes in pancreatic β -cell fractional area and β -cell turnover in human pregnancy. *Diabetologia* **53**, 2167–2176 (2010).
- Rieck, S. & Kaestner, K. H. Expansion of β -cell mass in response to pregnancy. *Trends Endocrinol. Metab.* **21**, 151–158 (2010).
- Ebert, M. *et al.* Induction of platelet-derived growth factor A and B chains and overexpression of their receptors in human pancreatic cancer. *Int. J. Cancer* **62**, 529–535 (1995).
- Nyblom, H. K. *et al.* Apoptotic, regenerative, and immune-related signaling in human islets from type 2 diabetes individuals. *J. Proteome Res.* **8**, 5650–5656 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Bhushan, A. Stewart, A. Powers, P. Beachy, M. White, X. Chen and X. Li for helpful discussions and advice, A. Tarakhovskiy, P. Herrera and M. Hara for mice, A. Powers, A. Thompson and S. Bryant for human islet sample procurement, and members of the S.K.K. laboratory for comments on the manuscript. H.C. was supported by the NIH Ruth L. Kirschstein NRSA/Stanford Regenerative Medicine Training Program. J.S. was supported by NIH-NCI R01 CA114102. H.S. was supported by the Stem Cell Network NRW and Deutsche Krebshilfe. Work in the S.K.K. laboratory was supported by a gift from the Dewey family fund, and grants from the Juvenile Diabetes Research Foundation, Snyder Foundation, Stinehart Foundation, the NIH Beta Cell Biology Consortium (U01 DK89532 to S.K.K. and U01 DK89572 to A. Powers) and by the Howard Hughes Medical Institute (HHMI). S.K.K. is an Investigator of the HHMI.

Author Contributions H.C., X.G., Y.L. and J.W. performed experiments. H.C., S.E.W., J.S. and H.S. generated mice. R.B. isolated human islets. H.C. and S.K.K. conceived the project, generated hypotheses, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K.K. (seungkim@stanford.edu).

METHODS

Animals. Transgenic mice harbouring a floxed *Pdgfra* allele⁴⁰ encoding *Pdgfr-α* were purchased from the Jackson Laboratory. This strain of mice was intercrossed with *RIP-Cre* mice⁴¹ expressing Cre recombinase from rat insulin 2 gene promoter elements to generate *RIP-Cre, Pdgfra^{fl/fl}* mice (designated as β Pdgfr α KO) and their littermate controls (including *RIP-Cre, Pdgfra^{fl/+}* mice designated as β Pdgfr α Het) on a mixed 129/Sv; C57BL/6 genetic background.

R26^{hPDGFRαPM} transgenic mice that harbour a mutated human *PDGFRA*^{D842V} allele encoding a ligand-independent activated receptor targeted immediately after a *loxP*-flanked transcriptional stop sequence at the *ROSA26* locus have been described previously²². This strain of mice was intercrossed with *RIP-Cre* mice to generate *RIP-Cre; R26-PDGFRα^{D842V}* mice (designated as β PDGFR α Tg) together with their littermate controls on a mixed 129/Sv; C57BL/6 genetic background. Mice with specific expression of mutant human *PDGFRA*^{D842V}, but lacking *Ezh2* in β -cells (designated as β Ezh2KO-R α Tg), were generated from subsequent intercrosses between β Ezh2KO¹³ and β PDGFR α Tg mice.

Tamoxifen-inducible, compound triple-mutant mice lacking retinoblastoma pocket proteins Rb, p130 and p107 (*ROSA26-CreER^{T2}; Rb^{fl/fl}; p130^{fl/fl}; p107^{-/-}*, designated as RbTriKO) were produced in the Sage group (Stanford University). RbTriKO mice harbour a germline-deleted *p107* null allele (*p107^{-/-}*), and *loxP*-flanked *Rb* and *p130* conditional alleles which can be excised by tamoxifen-sensitive CreER^{T2} recombinase targeted into the *ROSA26* locus²⁸. Littermate mice (*Rb^{fl/fl}; p130^{fl/fl}; p107^{-/-}*) lacking *ROSA26-CreER^{T2}* were used as controls. To delete *Rb* and *p130* alleles, RbTriKO and littermate control mice aged 3–5 months received intraperitoneal injections of tamoxifen (Sigma, 1.5 mg per mouse per day) dissolved in ethanol/corn oil (Sigma) on five consecutive days.

MIP-EGFP mice⁴² were obtained from the laboratory of M. Hara. C57BL/6 mice were purchased from the National Cancer Institute (NCI). All mutant mice used in this study were genotyped by PCR of tail genomic DNA for the human *PDGFRA*^{D842V} allele, mouse *Pdgfra^{fl}, Ezh2^{fl}, Rb^{fl}, p130^{fl}, p107^{-/-}* alleles, and transgenes encoding EGFP, *RIP-Cre* or *ROSA-CreER^{T2}* as described previously^{13,22,28,40–42}. Mice used in this study were age- and gender-matched littermates including both sexes. All mice were housed in the animal facility of Stanford University on a 12-h light/dark cycle with *ad libitum* access to water and normal chow except when otherwise indicated. All animal experiments and methods were approved by the Institutional Animal Care and Use Committee of Stanford University.

Physiological studies. We performed glucose physiology studies as previously reported¹³. Briefly, at the ages or times indicated in the text and figures, or after tamoxifen administration, we monitored body weight and blood glucose levels in *ad libitum* fed mice, or after 16 h overnight fasting, or 2 h after re-feeding (2 h postprandial). We drew blood from tail vein punctures or from the aorta when mice were killed, for measuring blood glucose levels using a Contour glucose meter (Bayer), or for measuring plasma insulin levels by mouse insulin ELISA kits (ALPCO Diagnostics).

Glucose tolerance tests were performed on mice after a 16 h overnight fast, and the blood glucose levels were determined immediately before (0) and 5, 10, 20, 30, 45, 60, 90 and 120 min after intraperitoneal injection of glucose (2 g kg⁻¹ body weight). For measurement of plasma insulin levels during these glucose tolerance tests, mice were injected with glucose at a dose of 3 g kg⁻¹ body weight, and we collected tail vein blood at 0, 15 and 45 min after glucose injection followed by mouse insulin ELISA (ALPCO diagnostics). For insulin tolerance tests, mice were fasted for 6 h then received intraperitoneal injection of 0.75 units kg⁻¹ of body weight of insulin (Sigma); glucose levels at 0, 10, 20, 30, 40, 60 and 90 min after injection were then measured.

For detecting *in vivo* β -cell proliferation by BrdU incorporation, mice were either fed with water containing BrdU (1 mg ml⁻¹) or intraperitoneally injected with BrdU/PBS solution (50 mg kg⁻¹ body weight) for specified schedules as follows: β Pdgfr α KO mice, 2 days of intraperitoneal injection of BrdU/PBS; β Pdgfr α KO mice with STZ treatment, 1 week of BrdU water feeding; β PDGFR α Tg or β Ezh2KO-R α Tg mouse studies, 6–7 days of BrdU water feeding; RbTriKO mice, 4 days of BrdU water feeding. After BrdU chase, mouse pancreases were collected and subjected to immunohistology followed by morphometric analyses described later.

STZ-induced diabetes. For measuring islet *Pdgfra* and *Pdgfrb* induction during STZ-induced diabetes, male wild-type C57BL/6 mice, aged 8 weeks, were injected intraperitoneally with 100 mg kg⁻¹ body weight of freshly dissolved streptozotocin (STZ, Sigma) in 0.1 mol l⁻¹ sodium citrate (pH 4.5) or with the same volume of sodium citrate (Vehicle). Six days later, islets or pancreases were recovered from the mice for gene expression analyses by real-time RT-PCR or immunohistology studies. To test the roles of *Pdgfra* in islet regeneration in STZ diabetes, β Pdgfr α KO or control mice (including β Pdgfr α Het), aged 7–8 weeks, were treated with a single intraperitoneal STZ injection at a dose of 150 mg kg⁻¹ body weight. After injection, blood glucose values were measured on day 3 and day 6 in the first week, and

thereafter once per week. After 2 weeks of STZ treatment, subsets of β Pdgfr α KO and control littermates were fed with BrdU-containing drinking water (1 mg ml⁻¹) for 1 week for BrdU incorporation analyses. Three days, 21 days and 58 days after STZ treatment, subsets of mice were killed, and blood samples and pancreases were harvested for measurement of plasma insulin levels by ELISA, and for pancreatic β -cell proliferation and mass quantification following immunostaining of insulin and BrdU.

Islet isolation and culture studies. Mouse pancreatic islets were purified as described previously¹³ and used immediately for assays described in the text. To study *Ezh2* induction in cultured islets exposed to mitogens, freshly isolated islets from 2–3-week old, and 7–12-month old C57BL/6 mice were recovered and equilibrated overnight at 37 °C, 5% CO₂ in RPMI-1640 medium supplemented with 0.5% fetal bovine serum (FBS), 0.2% bovine albumin (BSA) and 2% penicillin/streptomycin. Afterwards, islets were distributed into microplates or dishes, and treated with recombinant human PDGF-AA, prolactin or insulin (Sigma), with or without addition of pharmacological inhibitors, including Sunitinib (Selleck), Vargatef (Selleck), U0126 (Sigma), LY294002 (Sigma) or U-73122 (Sigma) at the indicated concentrations. Two days later, islets were harvested for RNA or protein extraction for real-time RT-PCR or western blot analyses. To detect phosphorylation of *Pdgfr-α* by western blot, islets were exposed to PDGF-AA (with or without indicated pharmacological inhibitors) for 90 min before islet harvesting and protein extraction. To detect *in vitro* islet β -cell proliferation stimulated by PDGF-AA, a final concentration of 50 μM BrdU was introduced into each culture dish 24 h before harvest, followed by standard histological processing. Immunostaining of BrdU and insulin was performed on processed islets, and β -cell proliferation rate was determined by quantifying percentage of insulin⁺ cells with BrdU⁺. Culture experiments of human islets were performed using a similar procedure described later.

β -Cell purification by FACS. Freshly purified islets from MIP-GFP mice at 2 weeks or 5 months of age were pooled, and dispersed into a suspension of single cells by incubation with 0.05% trypsin/0.53 mM EDTA solution at 37 °C for 10 min as described previously^{37,43}. After dissociation, cells were washed three times in PBS containing 2% FBS, and re-suspended in FACS buffer (HBSS/0.2% BSA/1% HEPES) containing propidium iodide (to exclude dead cells). FACS was performed on a FACS AriaII (BD Biosciences), and MIP-GFP⁺ β -cells with >95% purity were harvested and immediately lysed for RNA purification and real-time RT-PCR analysis.

Real-time RT-PCR and western blot analysis. Total RNA from freshly isolated or cultured islets after indicated treatments were purified using the Absolutely RNA miniprep purification kit (Stratagene) according to the manufacturer's instructions. RNA concentration was measured with a RiboGreen RNA quantification assay (Invitrogen). One-step quantitative RT-PCR was performed and analysed using an ABI Prism 7300 detection system (Applied Biosystems) with TaqMan one-step RT-PCR Master Mix Reagents and appropriate amounts (10–100 ng) of islet total RNA as the template. We calculated the ratio of mRNA for the gene of interest to the amount of internal control mRNA of peptidylprolyl isomerase A (cyclophilin A, PPIA), and then normalized the ratio for each gene to its median. Primer and probe sequences are listed in Supplementary Tables 1 and 2.

For western blots, total islet protein was prepared from freshly isolated or cultured islets after indicated treatments as previously described¹³. Equal amounts of protein were resolved on SDS-PAGE and transferred to polyvinylidene fluoride membranes (Amersham Pharmacia) for immunoblotting with specific antibodies, including rabbit polyclonal anti-human *Ezh2* (1:1,000, Millipore), rabbit polyclonal anti-phospho-PDGFR- α (Tyr 720) (1:1,000, Santa Cruz Biotechnologies), rabbit monoclonal anti-PDGFR- α (1:1,000, Cell Signaling), mouse monoclonal anti- β -actin (1:4,000, Sigma), rabbit monoclonal anti-Akt (1:1,500, Cell Signaling), rabbit monoclonal anti-phospho-Ark (Ser 473) (1:1,000, Cell Signaling), rabbit monoclonal anti-p44/42 MAPK (Erk1/2) (1:1,500, Cell Signaling), rabbit monoclonal anti-phospho-p44/42 MAPK (Erk1/2) (Thr 202/Tyr 204) (1:1,000, Cell Signaling), rabbit monoclonal anti-PLC γ 1 (1:1,000, Cell Signaling) and rabbit monoclonal anti-phospho-PLC γ 1 (Tyr 783) (1:1,000, Cell Signaling). Signal was visualized using ECL detection (Amersham Pharmacia) on Kodak film after further incubation with HRP-conjugated secondary antibodies. For phospho-PDGFR- α (Tyr 720) and PDGFR- α detection, signal was amplified by incubating the membrane with biotinylated anti-rabbit-IgG (1:2,000), followed with an incubation with HRP-avidin (1:2,000, Vector Laboratories) and then ECL visualization.

Histology, immunofluorescence and immunohistochemistry. We performed standard histological paraformaldehyde fixation, paraffin-embedding and immunostaining protocol as previously described¹³. Briefly, immunohistochemical analysis was performed on 5-μm-thick sections of pancreatic tissues or islet sections after antigen retrieval (DAKO) and using the following primary antibodies: guinea pig anti-insulin (1:400, Sigma), mouse anti-glucagon (1:200, Sigma), rabbit anti-somatostatin (1:200, Dakocytomation), rabbit anti-pancreatic polypeptide

(1:200, Dakocytomation), rabbit anti-Ezh2 (1:100, Epigentek), rabbit anti-PDGFR- α (1:50, Novus Biologicals), sheep anti-PDGFR- β (1:100, Sigma), rabbit anti-phospho-PDGFR- α (Tyr 720) (1:50, Santa Cruz Biotechnologies), rabbit anti-phospho-p44/42 MAPK (Erk1/2) (Thr 202/Tyr 204) (1:50, Cell Signaling), rabbit anti-phospho-Rb (Ser 780) (1:100, Cell Signaling), rabbit anti-cyclin D1 (1:50, Cell Signaling), rabbit anti-cyclin D2 (1:50, Cell Signaling), mouse anti-PDX1 (1:50; gift from H. Edlund), mouse monoclonal anti-Ki67 (1:100, Novocastra) and mouse monoclonal anti-BrdU (1:100, Sigma). We detected immune complexes with secondary antibodies conjugated with either Cy3, fluorescein isothiocyanate (Jackson ImmunoResearch) or horseradish peroxidase (Vector Laboratories). After staining, images were directly collected on an AxioM1 microscope equipped with a CCD digital camera (Carl Zeiss), or on a Leica SP2 inverted confocal microscope.

For measuring pancreatic β -cell mass or BrdU by morphometry, three to six mice for each group were analysed. At least five sections separated by more than 300 μ m (pancreases) or 50 μ m (cultured islets) were immunostained and assessed for each sample. Images were analysed with an ImagePro program by observers blinded to genotype, and pancreatic β -cell mass and percentage of BrdU⁺ β -cells were calculated as described previously¹³.

Pancreatic islet ChIP analysis. ChIPs were performed using a chromatin immunoprecipitation kit from Millipore as described previously¹³. Briefly, freshly isolated mouse pancreatic islets or cultured human islets after the indicated treatments were pooled, and fixed with 2% formaldehyde for cross-linking for 20 min at room temperature (25 °C). After washing, islets were dissolved in SDS lysis buffer containing a proteinase inhibitor cocktail followed by sonication to shear the chromatin. Precleared chromatin from 150 to 300 islets was used for each ChIP sample with incubation of 1 to 5 μ g of the appropriate antibodies at 4 °C overnight. The antibodies used in the ChIP assays included mouse anti-E2F1 (Millipore), rabbit anti-E2F4 (Santa Cruz Biotechnologies), rabbit or mouse polyclonal IgG (Millipore). After incubation, the immunoprecipitated chromatin DNA was harvested, cross-link reversed, and purified. After measurement of DNA concentration by PicoGreen DNA assay (Invitrogen), equivalent amounts of chromatin DNA from every sample was quantified by real-time qPCR in ABI Prism 7300 detection system (Applied Biosystems). The sequences of the PCR primers and probes are listed in Supplementary Table 3.

Human pancreas and islet studies. Human pancreases or islets from organ donors were procured by arrangement with the National Disease Resource Interchange and the University of Alabama, Birmingham with appropriate consent.

Institutional review board approval for research use of tissue was obtained from Stanford University School of Medicine. For pancreas studies, fresh human pancreases were processed in our laboratory into paraffin sections using standard histological protocols. Four pancreases from juvenile donors (1, 2, 3, 7 years old) and four pancreases from adult donors (22, 34, 55, 56 years old) were used for immunohistology studies here. Similar immunohistology methods and antibodies as those used for mouse pancreas tissues were used to detect PDGFR- α , phospho-ERK1/2 (pERK1/2), phospho-RB (Ser 780) (pRB(Ser780), EZH2 and insulin on human pancreas sections. Experiments were repeated on multiple pancreas sections for every donor pancreas and similar results were obtained.

For islet studies, human islets were isolated either at Pittsburgh University Medical Center (R.B.), or the University of Alabama, Birmingham (S. Bryant and A. Thompson). Eight independent human islet preparations from juvenile donors at the age of 8 months, 1.0, 1.5, 2, 3, 4, 5 and 6 years, and three adult islet preparations from donors at the age of 39, 48 and 56 years were used in this study. After isolation, fresh islets were shipped to our laboratory, and were hand-picked and transferred to RPMI-1640 medium supplemented with 0.5% FBS, 0.2% BSA and 2% penicillin/streptomycin for an overnight recovery before experiments. Similar to the experiments performed with mouse islets, equilibrated human islets were exposed to recombinant human PDGF-AA, without or with addition of pharmacological inhibitors, including Sunitinib (Selleck) and U0126 (Sigma) for 2 days, followed by islet RNA purification for real-time RT-PCR analyses of gene expression. To detect *in vitro* human islet β -cell proliferation stimulated by PDGF-AA, human islets were exposed to 50 μ M BrdU for 24 h. Afterwards, human islets were processed in a similar way as in mouse islet studies for immunostaining of insulin, glucagon, PDX1 and BrdU. β -Cell proliferation rate was determined by quantifying percentage of insulin⁺ cells with BrdU⁺. More than 50 islets per group were counted for determining cultured human islet β -cell proliferation rate.

40. Tallquist, M. D. & Soriano, P. Cell autonomous requirement for PDGFR α in populations of cranial and cardiac neural crest cells. *Development* **130**, 507–518 (2003).
41. Herrera, P. L. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* **127**, 2317–2322 (2000).
42. Hara, M. *et al.* Transgenic mice with green fluorescent protein-labeled pancreatic β -cells. *Am. J. Physiol. Endocrinol. Metab.* **284**, E177–E183 (2003).
43. Sugiyama, T. *et al.* Conserved markers of fetal pancreatic epithelium permit prospective isolation of islet progenitor cells by FACS. *Proc. Natl Acad. Sci. USA* **104**, 175–180 (2007).

Macrophage skewing by *Phd2* haploinsufficiency prevents ischaemia by inducing arteriogenesis

Yukiji Takeda^{1,2,3,4}, Sandra Costa^{1,2,5*}, Estelle Delamarre^{1,2*}, Carmen Roncal^{1,2,3,4,6*}, Rodrigo Leite de Oliveira^{1,2,3,4}, Mario Leonardo Squadrito^{7,8}, Veronica Finisguerra^{1,2}, Sofie Deschoemaeker^{1,2}, Françoise Bruyère^{3,4}, Mathias Wenes^{1,2}, Alexander Hamm^{1,2}, Jens Serneels^{1,2}, Julie Magat⁹, Tapan Bhattacharyya^{10†}, Andrey Anisimov¹¹, Benedicte F. Jordan⁹, Kari Alitalo¹¹, Patrick Maxwell¹⁰, Bernard Gallez⁹, Zhen W. Zhuang¹², Yoshihiko Saito¹³, Michael Simons¹², Michele De Palma^{7†} & Massimiliano Mazzone^{1,2}

PHD2 serves as an oxygen sensor that rescues blood supply by regulating vessel formation and shape in case of oxygen shortage^{1–5}. However, it is unknown whether PHD2 can influence arteriogenesis. Here we studied the role of PHD2 in collateral artery growth by using hindlimb ischaemia as a model, a process that compensates for the lack of blood flow in case of major arterial occlusion^{6–8}. We show that *Phd2* (also known as *Egln1*) haploinsufficient (*Phd2*^{+/-}) mice displayed preformed collateral arteries that preserved limb perfusion and prevented tissue necrosis in ischaemia. Improved arteriogenesis in *Phd2*^{+/-} mice was due to an expansion of tissue-resident, M2-like macrophages^{9,10} and their increased release of arteriogenic factors, leading to enhanced smooth muscle cell (SMC) recruitment and growth. Both chronic and acute deletion of one *Phd2* allele in macrophages was sufficient to skew their polarization towards a pro-arteriogenic phenotype. Mechanistically, collateral vessel preconditioning relied on the activation of canonical NF-κB pathway in *Phd2*^{+/-} macrophages. These results unravel how PHD2 regulates arteriogenesis and artery homeostasis by controlling a specific differentiation state in macrophages and suggest new treatment options for ischaemic disorders.

To understand whether partial loss of PHD2 enhances perfusion of ischaemic tissues, we subjected mice to femoral artery ligation, a procedure that reduces perfusion of the lower limb, causing ischaemia in the calf (that is, the crural muscle). After ligation, *Phd2*^{+/-} mice showed a milder drop in perfusion and oxygen tension with reduced hypoxia in the crural muscle compared to wild-type (WT) mice (Fig. 1a–g). Ischaemia promotes oxidative stress early and angiogenesis as a later response^{6,11}. Oxidative stress (12 h post-ligation) and capillarization (14 days post-ligation) were both increased in the crural muscle of WT, but not *Phd2*^{+/-} mice (Supplementary Fig. 2a–g). As a consequence of preserved blood flow, *Phd2*^{+/-} crural muscles showed reduced ischaemic necrosis and increased viability (Fig. 1h–j and Supplementary Fig. 2h–j). WT mice showed signs of muscle regeneration that were absent in *Phd2*^{+/-} crural muscles (Supplementary Fig. 2k–m). The protection against ischaemic damage improved physical endurance of *Phd2*^{+/-} mice in ischaemia (Fig. 1k), although both genotypes had similar running capacity at baseline (Supplementary Fig. 2n).

Because *Phd2*^{+/-} mice were protected against ischaemia already 12 h post-ligation, we proposed that they could tolerate ischaemic insults better due to increased collaterals at baseline¹². Macroscopic inspection of the upper limb, that is, the thigh, and histological analysis

of the adductor (in the inner thigh, where collaterals grow) after gelatin-bismuth angiographies showed about twice higher numbers and area of bismuth-positive collaterals in non-ligated *Phd2*^{+/-} versus WT mice (Fig. 1l–q and Supplementary Fig. 3a,b). Also micro-computed tomography scans and X-ray radiographies showed higher numbers of large vessels (>200 μm in diameter) in *Phd2*^{+/-} than WT thighs at baseline (Fig. 1r–t and Supplementary Fig. 3c,d), whereas numbers of smaller vessels (<200 μm in diameter) and capillaries were comparable in both genotypes (Supplementary Fig. 3e–g). Both the total area and numbers of bismuth-positive collaterals were still higher in *Phd2*^{+/-} versus WT adductors 12 and 72 h post-ligation, a time-window when collateral remodelling just begins in WT mice (Fig. 1p,q and Supplementary Fig. 3a,b).

We also assessed whether *Phd2*^{+/-} mice were protected against myocardial ischaemia. Twenty-four hours after coronary artery ligation, desmin-negative area (a readout of cardiomyocyte death) was smaller in *Phd2*^{+/-} hearts (Supplementary Fig. 4a–c). Compared to WT, *Phd2*^{+/-} hearts showed higher perfusion in both infarcted and remote myocardium (Supplementary Fig. 4d–g). At baseline, density of large vessels, but not small vessels and capillaries, was higher in *Phd2*^{+/-} versus WT hearts (Supplementary Fig. 4h–l).

To increase blood flow in case of major arterial occlusion, collateral vessels undergo extensive remodelling (arteriogenesis) with thickening of the tunica media, consisting of α-smooth muscle actin (αSMA)-positive SMCs, and enlargement of vessel diameter⁸. Numbers and total area of αSMA⁺ collateral vessels were higher in *Phd2*^{+/-} adductors both at baseline and after ischaemia, whereas the mean area and thickness of the tunica media were higher only at baseline (Fig. 1u–b'). These data show that collateral vessels of *Phd2*^{+/-} mice at baseline were similar to those of WT mice after femoral artery ligation. This 'collateral vessel preconditioning' was protective against ischaemia.

Inflammatory cells and macrophages in particular are responsible for collateral vessel remodelling^{7,8}. Nevertheless, CD45⁺ leukocyte and F4/80⁺ macrophage infiltration of the adductors was similar at baseline and equally increased after ligation in both genotypes (Fig. 2a,b). We therefore analysed the phenotype of infiltrating macrophages and measured the density of M2-like, wound-healing/pro-angiogenic macrophages by their expression of the MRC1 mannose receptor^{9,10}. At baseline, F4/80⁺MRC1⁺ macrophages were 75% higher in *Phd2*^{+/-} versus WT adductors (Fig. 2c–e). Seventy-two hours after ligation, their numbers were increased by 95% in WT and only 50% in *Phd2*^{+/-} mice

¹Laboratory of Molecular Oncology and Angiogenesis, Vesalius Research Center, VIB, Leuven B-3000, Belgium. ²Laboratory of Molecular Oncology and Angiogenesis, Vesalius Research Center, K. U. Leuven, Leuven B-3000, Belgium. ³Laboratory of Angiogenesis and Neurovascular link, Vesalius Research Center, VIB, Leuven B-3000, Belgium. ⁴Laboratory of Angiogenesis and Neurovascular link, Vesalius Research Center, K. U. Leuven, Leuven B-3000, Belgium. ⁵Life and Health Sciences Research Institute, Minho University, 4710-057 Braga, Portugal. ⁶Atherosclerosis Research Laboratory, CIMA-University of Navarra, 31008 Pamplona, Spain. ⁷Angiogenesis and Tumor Targeting Unit and HSR-TIGET, San Raffaele Institute, 20132 Milan, Italy. ⁸Vita-Salute University, 20132 Milan, Italy. ⁹Biomedical Magnetic Resonance Unit, Université Catholique de Louvain, Brussels B-1200, Belgium. ¹⁰Rayne Institute, University College London, London WC1E 6JF, UK. ¹¹Molecular/Cancer Biology Laboratory, Research Programs Unit and Institute for Molecular Medicine, Biomedicum Helsinki, 00014 Helsinki, Finland. ¹²Cardiovascular Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA. ¹³The First Department of Internal Medicine, Nara Medical University, 634-8522 Nara, Japan. †Present addresses: Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK (T.P.); The Swiss Institute for Experimental Cancer Research (ISREC), Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, CH-1015, Switzerland (M.D.P.).

*These authors contributed equally to this work.

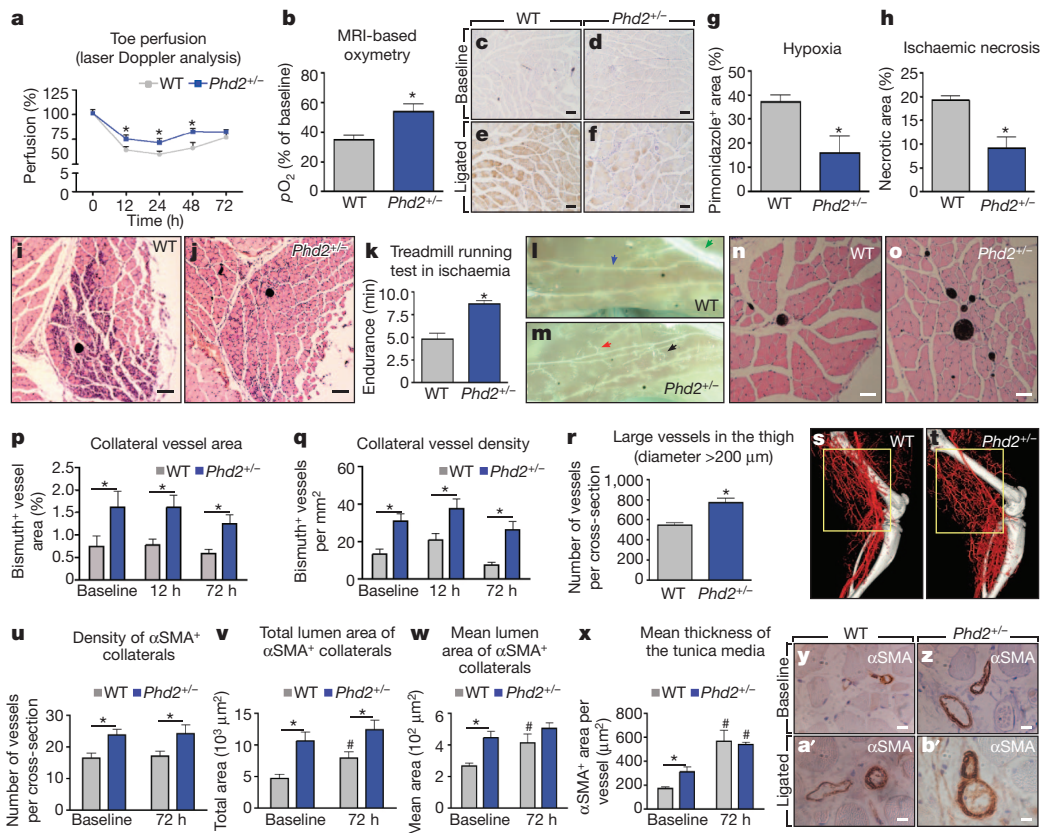


Figure 1 | *Phd2*^{+/-} mice are pre-adapted to ischaemia. **a**, Laser Doppler analysis 12, 24, 48 and 72 h post-ligation. **b**, Magnetic resonance imaging (MRI) oxymetry in crural muscles 12 h post-ligation. **c–g**, Hypoxic area quantification (**e, f**); hypoxia was undetectable at baseline (**c, d**). Scale bars, 200 μm. **g**, Pimonidazole⁺ area 12 h post-ligation. **h–j**, Necrotic area quantification (**i, j**) on haematoxylin- and eosin-stained sections of crural muscles 72 h post-ligation (**i, j**). Scale bars, 100 μm. **k**, Treadmill running test 12 h post-ligation. **l, m**, Macroscopic view of adductors after gelatin-bismuth-based angiographies at baseline. Collateral vessels: primary (blue arrow), secondary (red arrow), and

tertiary (black arrow). Femoral artery (green arrow). **n, o**, Haematoxylin and eosin staining of adductors at baseline after angiographies; bismuth⁺ collaterals appear black. Scale bars, 50 μm. **p, q**, Collateral vessel area (**p**) and density (**q**) represented in **n, o, r**. Quantification of large vessels (>200 μm in diameter) in the thigh at baseline after micro-computed tomography angiograms. **s, t**, Representative micro-computed tomography micrographs of the thigh (yellow frame). **u–x**, Morphological analysis on αSMA-stained sections of adductors at baseline and 72 h post-ligation, as represented in **y–b'**. Scale bars, 10 μm. All graphs show mean ± s.e.m. All experiments, *n* ≥ 5. *, *P* < 0.05 towards WT. #, *P* < 0.05 towards baseline.

(Fig. 2c, f, g). Gene-profiling of *Phd2*^{+/-} peritoneal macrophages (pMΦ) showed higher expression of M2-type genes^{9,10,13}, including *Tek* (also known as *Tie2*), *Arg1*, *Cxcr4*, *Ccr2*, *Hgf*, *Pdgfb*, *Fizz*, *Nrp1*, *Mmp2*, *Cxcl12* (also known as *Sdf1*) and *Tgfb*, than WT pMΦ (Fig. 2h). Conversely, several proinflammatory or anti-angiogenic (that is, M1-type) molecules were downregulated in *Phd2*^{+/-} macrophages; these included *Il1b*, *Il6*, *Nos2* and *Il12* (Fig. 2h). Similarly, *Phd2*^{+/-} macrophages sorted from the adductor expressed higher levels of *Pdgfb*, *Sdf1*, *Tie2*, *Mmp2* and *Nrp1* at baseline (Fig. 2i). Seventy-two hours post-ligation, the expression level of these markers was similar in *Phd2*^{+/-} and WT tissue macrophages (Fig. 2i). Expression of these genes was comparable in WT and *Phd2*^{+/-} endothelial cells isolated from adductors at baseline or in ischaemia (Supplementary Table 1). Noteworthy, the basal level of *Phd2* in *Phd2*^{+/-} macrophages was half of that in WT macrophages and did not change in ischaemia (Fig. 2i). Conversely, *Phd2* expression in WT macrophages was reduced by ~50% in ischaemia and thus reached the same level as in *Phd2*^{+/-} macrophages (Fig. 2i). Thus, *Phd2*^{+/-} macrophages showed a unique and cell-specific gene signature, which was reminiscent, at least in part, of M2-polarized macrophages and of WT macrophages in ischaemia.

We therefore assessed whether WT and *Phd2*^{+/-} macrophages affect the behaviour of endothelial cells and SMCs, the two main cellular components of arteries. Soluble factors released by *Phd2*^{+/-} macrophages strongly increased migration and proliferation of SMCs, but not endothelial cells, probably because the latter were already highly responsive to WT macrophages (Fig. 2j–n and

Supplementary Fig. 5a–e). Consistently, SMCs exposed to conditioned medium from *Phd2*^{+/-} macrophages showed reduced levels of calponin-1, Sm22, smoothelin, NMHC-B and αSMA (Fig. 2o–s), indicating enhanced proliferation^{14,15}. *In vitro* knockdown of both *Sdf1* and *Pdgfb*, known to stimulate SMC recruitment and proliferation^{16,17}, abolished the enhanced response of SMCs to *Phd2*^{+/-} macrophages, although inhibition of either factor was also very effective (Supplementary Fig. 5f, g and Supplementary Note 1). Overall, these data showed that lower levels of PHD2 pre-adapt macrophages to ischaemia by skewing them towards an M2-like phenotype, which promotes SMC recruitment and growth.

We then investigated whether reduced levels of PHD2 in macrophages promoted collateral vessel preconditioning. Myeloid-cell specific *Phd2* haploinsufficiency (*Phd2*^{LysCre/lox/WT}) increased numbers and area of collateral branch arteries, thus conferring protection against ischaemic necrosis and enhancing running capacity in ischaemia (Fig. 3a–i and Supplementary Fig. 6a). In myeloid-cell-specific *Phd2*-null mice (*Phd2*^{LysCre/lox/lox}), arterIALIZATION, ischaemic necrosis and physical endurance were unchanged (Fig. 3a–i and Supplementary Fig. 6a), probably because of the compensatory activity of PHD3, another PHD family member (see below). We also transplanted WT or *Phd2*^{+/-} (hereafter HE for 'heterozygous') bone marrow (BM) cells into irradiated WT (referred to as WT→WT and HE→WT mice, respectively) or *Phd2*^{+/-} (WT→HE and HE→HE mice, respectively) recipients (Supplementary Note 2). Compared to WT→WT, collateral vessel density and area were higher in HE→WT and HE→HE but

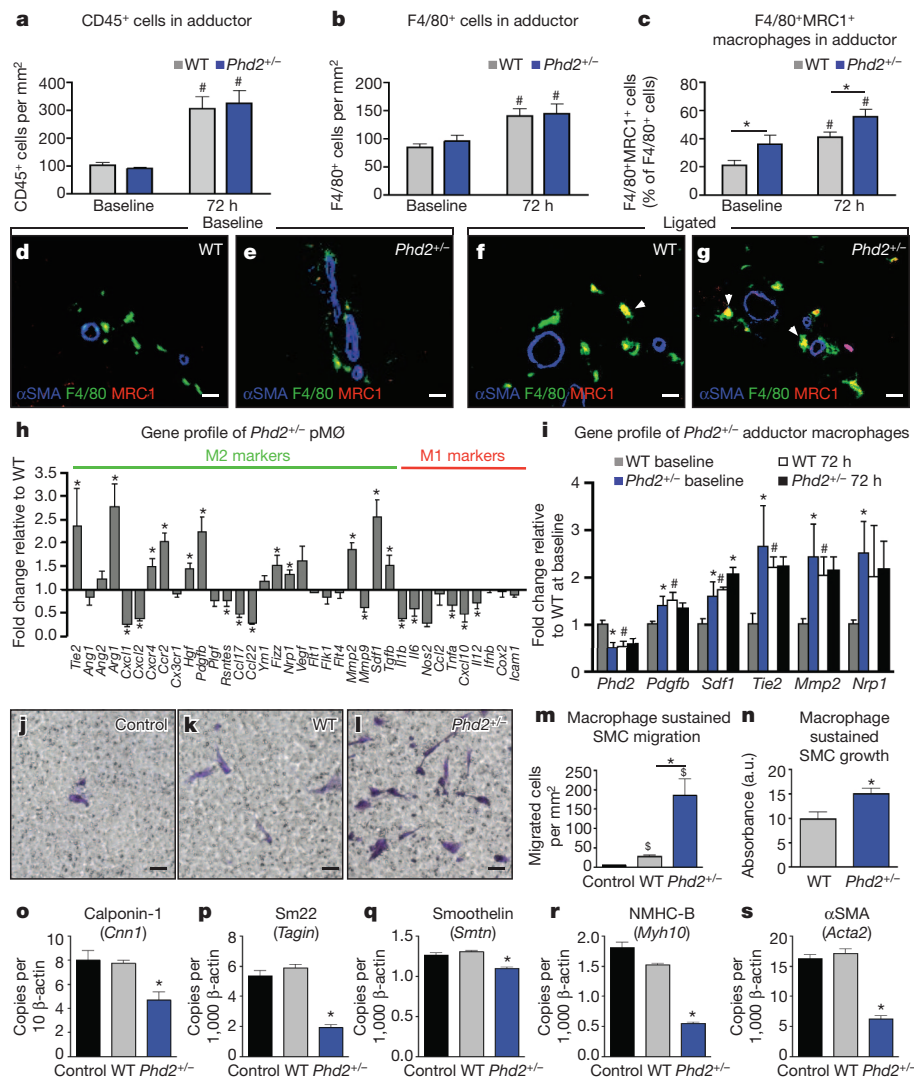
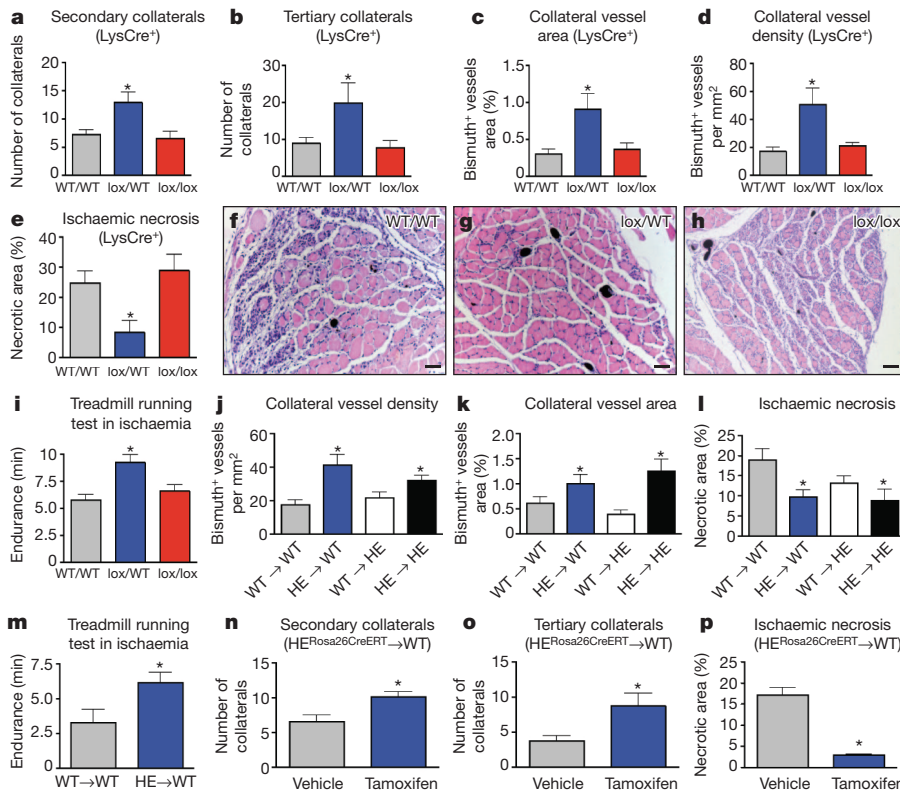


Figure 2 | *Phd2*^{+/-} macrophages display an M2-like phenotype. **a, b**, Quantification of CD45⁺ leukocyte and F4/80⁺ macrophage infiltration of adductors at baseline and 72 h post-ligation. **c**, Quantification of MRC1⁺ macrophages in adductors at baseline and 72 h post-ligation, represented in **d–g**. **d–g**, Micrographs of immunostainings for F4/80 (green), MRC1 (red) and αSMA (blue). Arrowheads (**f, g**) point to F4/80⁺MRC1⁺ cells. Scale bars, 20 μm. **h**, Gene profile by quantitative PCR of peritoneal macrophages (pMØ). **i**, Gene profile of F4/80⁺ macrophages sorted from adductors at baseline and 72 h post-ligation. **j–m**, Quantification (**m**) of crystal-violet-stained SMCs, migrated towards control medium (**j**), WT (**k**) or *Phd2*^{+/-} (**l**) macrophages. Scale bars, 50 μm. **n**, SMC growth in response to soluble factors released by WT or *Phd2*^{+/-} macrophages. **o–s**, Gene profile by quantitative PCR of SMCs exposed to control, WT macrophage-conditioned medium, or *Phd2*^{+/-} macrophage-conditioned medium. All bars show mean ± s.e.m. All experiments, *n* ≥ 5, **P* < 0.05 towards WT. #, *P* < 0.05 towards baseline. \$, *P* < 0.05 towards control medium.

similar in WT→HE mice, supporting the key role of BM-derived cells in enhancing collateralization and sustaining pre-existing arteries in *Phd2*^{+/-} mice (Fig. 3j, k). In accordance, HE→HE and HE→WT, but not WT→HE mice, were protected against ischaemic necrosis (Fig. 3l). The running capacity of HE→WT mice in ischaemia was twice as high as in WT→WT mice (Fig. 3m and Supplementary Fig. 6b). Deletion of one *Phd2* allele in haematopoietic cells, endothelial cells or SMCs confirmed that enhanced collateral vessel growth and maintenance were specifically conferred by *Phd2* haploinsufficiency in BM cells, but not endothelial cells or SMCs (Supplementary Tables 2 and 3). We also assessed the effect of acute deletion of *Phd2* in macrophages by transplanting BM cells from tamoxifen-inducible *Phd2*-haploinsufficient mice (*Phd2*^{Rosa26CreERT²/lox}/WT) into WT recipient mice (HE^{Rosa26CreERT²}→WT). Tamoxifen-induced deletion of one *Phd2* allele in BM cells increased collateral branches and protected against ischaemic necrosis when compared to vehicle (Fig. 3n–p and Supplementary Note 3). Thus, both chronic and acute deletion of one *Phd2* allele in myeloid cells was sufficient to induce pro-arteriogenic macrophages, leading to enhanced collateralization and prevention of ischaemia.

PHD oxygen sensors negatively regulate HIF accumulation and NF-κB activity^{1,11,18–22}. Compared to WT, HIF-1α and HIF-2α protein levels were respectively four and two times higher in *Phd2*-null macrophages, but unchanged in *Phd2*-haploinsufficient macrophages (Fig. 4a). Conversely, NF-κB activity was increased by 65% in *Phd2*-haploinsufficient macrophages but unaffected in *Phd2*-null macrophages (Fig. 4b). We proposed that PHD3 might compensate for the complete loss of *Phd2*

because *Phd3* transcripts were 12.2-fold higher in *Phd2*-null macrophages (Supplementary Fig. 7). Indeed, *Phd3* silencing resulted in a modest induction of NF-κB activity in WT and *Phd2*-haploinsufficient macrophages in contrast to a 70% increase in *Phd2*-null macrophages compared to their scramble controls (Fig. 4b and Supplementary Note 4). PHD2 hydroxylase function was necessary for NF-κB regulation because ectopic expression of a wild-type PHD2 (PHD2^{WT}) greatly blunted the activity of NF-κB luciferase induced by *Phd2* haploinsufficiency, whereas a hydroxylase-deficient PHD2 (PHD2^{H313A}) had no effect (Fig. 4c). NF-κB activation by TNF-α was still significantly stronger in *Phd2*^{+/-} macrophages (Fig. 4d). In contrast, basal and TNF-α-induced NF-κB activity were similar in WT and *Phd2*^{+/-} endothelial cells (Supplementary Fig. 8a). When measuring the nuclear accumulation of NF-κB subunits, we found that members of the canonical pathway, that is, p65 (RelA) and p50 (NF-κB1), were more abundant in *Phd2*^{+/-} than WT macrophages (Fig. 4e). Silencing of p65 or p50 blocked NF-κB hyperactivation in *Phd2*^{+/-} macrophages and the combined knockdown of both subunits restored NF-κB function back to WT levels (Fig. 4f and Supplementary Note 5), thus highlighting the prominent role of NF-κB p65/p50 heterodimers in *Phd2*^{+/-} macrophages. To evaluate the involvement of canonical NF-κB signaling in macrophage skewing by *Phd2* haploinsufficiency, we generated a myeloid-cell specific double transgenic strain, haploinsufficient for *Phd2* and null for *Ikkbb*, the gene encoding IKKβ, a positive regulator of canonical NF-κB pathway. Genetic disruption (or pharmacological inhibition) of canonical NF-κB pathway prevented the upregulation



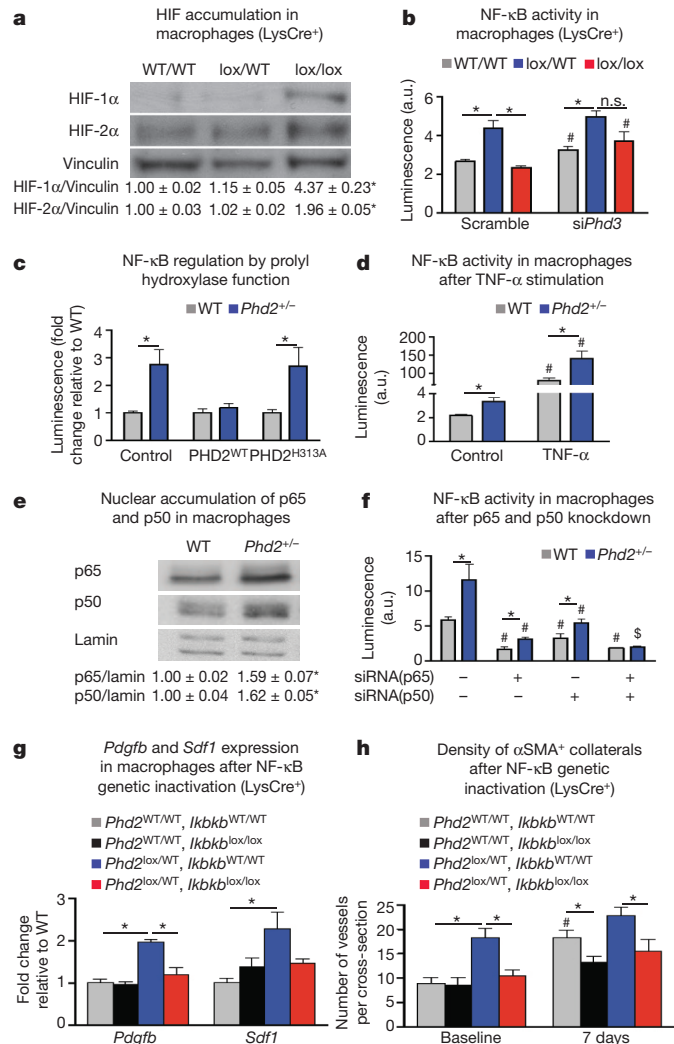
of *Pdgfb* and *Sdf1* in cultured *Phd2*-haplodeficient macrophages (Fig. 4g and Supplementary Fig. 8b). *In vivo*, gene inactivation of *Ikbkb* in myeloid cells abolished collateral vessel preconditioning conferred by *Phd2* haplodeficiency and greatly prevented ischaemia-induced arteriogenesis in WT mice (Fig. 4h). Thus, skewing of *Phd2*-haplodeficient macrophages towards a pro-arteriogenic phenotype relied on activation of canonical NF-κB pathway.

Specific macrophage differentiation states have been implicated in the promotion of angiogenesis during cancer and atherosclerosis progression^{9,10,13,23}. However, little is known of the significance of macrophage heterogeneity in arteriogenesis and its implications on ischaemic diseases. This study identifies a role of macrophage PHD2 in oxygen delivery by regulating arteriogenesis. We show that the phenotype of macrophages induced by reduced PHD2 levels not only favours the formation of new collateral branches, but is also important for collateral vessel homeostasis (Supplementary Note 6). In our model, ischaemia favours the accumulation of M2-like, pro-arteriogenic macrophages that fuel collateral vessel maturation in a NF-κB-dependent manner (Supplementary Fig. 1). *Phd2* haplodeficiency unleashes constitutive

Figure 4 | *Phd2*^{+/-} macrophages trigger arteriogenesis via canonical NF-κB pathway. **a**, Western blot for HIF-1α and HIF-2α in WT (WT/WT), *Phd2*-haplodeficient (lox/WT) and *Phd2*-null (lox/lox) macrophages. Vinculin was used as loading control. Numbers represent densitometric fold change towards WT/WT. **b**, NF-κB activity in macrophages after *Phd3* silencing (si*Phd3*; n.s., not significant). **c**, NF-κB activity in *Phd2*^{+/-} macrophages upon ectopic expression of wild-type PHD2 (PHD2^{WT}) or hydroxylase-deficient PHD2 (PHD2^{H313A}). **d**, NF-κB activity in macrophages at baseline and upon TNF-α stimulation. **e**, Western blot for nuclear p65 and p50 in WT and *Phd2*^{+/-} macrophages. Numbers represent densitometric fold change towards WT. **f**, NF-κB activity in macrophages after silencing of p65, p50, or combination of both. **g**, *Pdgfb* and *Sdf1* expression (quantitative PCR) after genetic inactivation of *Ikbkb* in *Phd2*-haplodeficient pMØ (*Phd2*^{lox/WT}, *Ikbkb*^{lox/lox}). **h**, Quantification of αSMA⁺ collaterals in myeloid-cell-specific *Phd2*-haplodeficient and *Ikbkb*-null mice at baseline and 7 days post-ligation. All bars and values show mean ± s.e.m. All experiments, *n* ≥ 4. *, *P* < 0.05 towards wt/wt in **a**, **b**, **g**, **h**, or towards WT in **c**–**f**. #, *P* < 0.05 towards scramble in **b**, **f**, or towards baseline in **d**, **h**. \$, *P* < 0.05 towards scramble and either short interfering RNA alone.

Figure 3 | *Phd2*^{+/-} macrophages protect against ischaemia by inducing arteriogenesis.

a, b, Quantification of secondary (**a**) and tertiary (**b**) collateral branches in WT (*Phd2*^{LysCre};WT/WT; WT/WT), myeloid-cell specific *Phd2*-haplodeficient (*Phd2*^{LysCre};lox/lox; lox/WT) or null mice (*Phd2*^{LysCre};lox/lox; lox/lox). **c, d**, Bismuth⁺ collateral vessel area (**c**) and density (**d**) of adductors at baseline. **e–h**, Necrotic area quantification (**e**) on haematoxylin- and eosin-stained sections of crural muscles 72 h post-ligation (**f–h**). Scale bars, 100 μm. **i**, Treadmill running test 12 h post-ligation. **j, k**, Bismuth⁺ collateral vessel density (**j**) and area (**k**) of non-occluded limbs 5 weeks after bone marrow (BM) transplantation. *Phd2*^{+/-} BM in WT and *Phd2*^{+/-} recipient mice (HE → WT and HE → HE, respectively); WT BM in WT and *Phd2*^{+/-} recipient mice (WT → WT and WT → HE). **l**, Ischaemic necrosis 72 h post-ligation. **m**, Treadmill running test 12 h post-ligation in WT → WT and HE → WT mice. **n–p**, Quantification of secondary (**n**) and tertiary (**o**) collateral vessels at baseline and of ischaemic necrosis 72 h post-ligation (**p**), following tamoxifen-induced deletion of one *Phd2* allele in BM cells of HE^{Rosa26CreERT} → WT mice. All bars show mean ± s.e.m. All experiments, *n* ≥ 4. *, *P* < 0.05 towards WT/WT and lox/lox in **a–i**, towards WT → WT in **j–m**, or towards vehicle in **n–p**.



NF- κ B signals that pre-adapt tissue-resident macrophages to ischaemia, accounting for the enhanced arteriogenesis at baseline and thus protection against ischaemic tissue demise. In particular, we show that NF- κ B activation in *Phd2*^{+/-} macrophages increases the production of SDF1 and PDGFB, which are concurrently required to prime SMC migration and growth *in vitro*. The relevance of SDF1 and PDGFB *in vivo* remains to be established, and other soluble factors may also contribute.

Repression of canonical NF- κ B pathway in macrophages promotes upregulation of M1-type and downregulation of M2-type genes^{24–26}. PHDs negatively regulate NF- κ B through either hydroxylase-dependent or -independent inactivation of IKK β in different cell contexts^{19–22}. We show here that *Phd2* haploinsufficiency results in hyperactivation of canonical NF- κ B pathway in macrophages and promotion of the M2-phenotype via accumulation of both p65 and p50 subunits, and that this regulation requires PHD hydroxylase function.

In summary, our study provides an insight on how the PHD2 oxygen sensor regulates arteriogenesis by modulating macrophage phenotype. The mechanism upstream to arteriogenic PHD2 downregulation in M2-like macrophages remains to be established. Nonetheless, our findings support the rationale for therapeutic inhibition of PHD2. Previous studies showed that unspecific inhibitors of PHD2 or silencing of PHD2 promote therapeutic revascularization against ischaemia^{2–5}. However, angiogenesis is a late response and organ function might be compromised before new blood vessel formation is achieved. In contrast, arteriogenesis takes place on pre-existing vascular shunts and our data suggest that either PHD2 inhibitors or cell-therapy-based strategies using PHD2 hypomorphic macrophages might be exploited as preventive medicine to promote collateral vascularization in patients at risk of limb or heart ischaemia, such as diabetic or hypercholesterolemic patients.

METHODS SUMMARY

129/S6 or Balb/c WT and *Phd2*^{+/-} mice (8–12 weeks old) were obtained from our mouse facility. *Phd2*^{+/-} and *Phd2* conditional knockout mice were obtained as previously described¹. To induce hindlimb ischaemia, unilateral or bilateral ligations of the femoral artery and vein and the cutaneous vessels branching from the caudal femoral artery side branch were performed without damaging the nervus femoralis²⁷. Oxygen tension (pO₂) in the lower limb was measured 12 h after femoral artery ligation by ¹⁹F magnetic resonance imaging oxymetry. Adductors and crural muscles were dissected, fixed in 2% paraformaldehyde, dehydrated, embedded in paraffin and sectioned at 7- μ m thickness for histology, immunostaining and morphometry analysis. Macrophages were either collected from the peritoneal cavity of the mice (pMO) or derived from BM precursors as described before²⁸. Balb/c WT and *Phd2*^{+/-} recipient mice were irradiated with 7.5 Gy. Subsequently, 5 \times 10⁶ BM cells from green fluorescent protein⁺ (GFP⁺) WT or GFP⁺ *Phd2*^{+/-} mice were injected intravenously via the tail vein. Femoral artery ligation, treadmill running test and bismuth angiography were performed 5 weeks after BM reconstitution. Full Methods and any associated references are available in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 September 2010; accepted 23 August 2011.

Published online 9 October 2011.

- Mazzone, M. *et al.* Heterozygous deficiency of PHD2 restores tumor oxygenation and inhibits metastasis via endothelial normalization. *Cell* **136**, 839–851 (2009).
- Milkiewicz, M., Pugh, C. W. & Egginton, S. Inhibition of endogenous HIF inactivation induces angiogenesis in ischaemic skeletal muscles of mice. *J. Physiol. (Lond.)* **560**, 21–26 (2004).
- Nangaku, M. *et al.* A novel class of prolyl hydroxylase inhibitors induces angiogenesis and exerts organ protection against ischemia. *Arterioscler. Thromb. Vasc. Biol.* **27**, 2548–2554 (2007).
- Huang, M. *et al.* Short hairpin RNA interference therapy for ischemic heart disease. *Circulation* **118**, S226–S233 (2008).
- Loinard, C. *et al.* Inhibition of prolyl hydroxylase domain proteins promotes therapeutic revascularization. *Circulation* **120**, 50–59 (2009).
- Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395 (2000).
- Simons, M. & Ware, J. A. Therapeutic angiogenesis in cardiovascular disease. *Nature Rev. Drug Discov.* **2**, 863–872 (2003).

- Schaper, W. Collateral circulation: past and present. *Basic Res. Cardiol.* **104**, 5–21 (2009).
- Mantovani, A. & Sica, A. Macrophages, innate immunity and cancer: balance, tolerance, and diversity. *Curr. Opin. Immunol.* **22**, 231–237 (2010).
- Nucera, S., Biziato, D. & De Palma, M. The interplay between macrophages and angiogenesis in development, tissue injury and regeneration. *Int. J. Dev. Biol.* **55**, 495–503 (2011).
- Aragónes, J. *et al.* Deficiency or inhibition of oxygen sensor Phd1 induces hypoxia tolerance by reprogramming basal metabolism. *Nature Genet.* **40**, 170–180 (2008).
- Helisch, A. *et al.* Impact of mouse strain differences in innate hindlimb collateral vasculature. *Arterioscler. Thromb. Vasc. Biol.* **26**, 520–526 (2006).
- Squadrito, M. L. & De Palma, M. Macrophage regulation of tumor angiogenesis: Implications for cancer therapy. *Mol. Aspects Med.* **32**, 123–145 (2011).
- Kumar, M. S. & Owens, G. K. Combinatorial control of smooth muscle-specific gene expression. *Arterioscler. Thromb. Vasc. Biol.* **23**, 737–747 (2003).
- Wolf, C. *et al.* Vascular remodeling and altered protein expression during growth of coronary collateral arteries. *J. Mol. Cell. Cardiol.* **30**, 2291–2305 (1998).
- Karshovska, E., Zagorac, D., Zerneck, A., Weber, C. & Schober, A. A small molecule CXCR4 antagonist inhibits neointima formation and smooth muscle progenitor cell mobilization after arterial injury. *J. Thromb. Haemost.* **6**, 1812–1815 (2008).
- Hellstrom, M., Kalen, M., Lindahl, P., Abramsson, A. & Betsholtz, C. Role of PDGF-B and PDGFR- β in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126**, 3047–3055 (1999).
- Jaakkola, P. *et al.* Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
- Chan, D. A. *et al.* Tumor vasculature is regulated by PHD2-mediated angiogenesis and bone marrow-derived cell recruitment. *Cancer Cell* **15**, 527–538 (2009).
- Xue, J. *et al.* Prolyl hydroxylase-3 is down-regulated in colorectal cancer cells and inhibits IKK β independent of hydroxylase activity. *Gastroenterology* **138**, 606–615 (2010).
- Cummins, E. P. *et al.* Prolyl hydroxylase-1 negatively regulates I κ B kinase- β , giving insight into hypoxia-induced NF κ B activity. *Proc. Natl Acad. Sci. USA* **103**, 18154–18159 (2006).
- Fu, J. & Taubman, M. B. Prolyl hydroxylase EGLN3 regulates skeletal myoblast differentiation through an NF- κ B-dependent pathway. *J. Biol. Chem.* **285**, 8927–8935 (2010).
- Mantovani, A., Garlanda, C. & Locati, M. Macrophage diversity and polarization in atherosclerosis: a question of balance. *Arterioscler. Thromb. Vasc. Biol.* **29**, 1419–1423 (2009).
- Fong, C. H. *et al.* An antiinflammatory role for IKK β through the inhibition of “classical” macrophage activation. *J. Exp. Med.* **205**, 1269–1276 (2008).
- Hagemann, T. *et al.* “Re-educating” tumor-associated macrophages by targeting NF- κ B. *J. Exp. Med.* **205**, 1261–1268 (2008).
- Greten, F. R. *et al.* NF- κ B is a negative regulator of IL-1 β secretion as revealed by genetic and pharmacological inhibition of IKK β . *Cell* **130**, 918–931 (2007).
- Luttun, A. *et al.* Revascularization of ischemic tissues by PlGF treatment, and inhibition of tumor angiogenesis, arthritis and atherosclerosis by anti-Fit1. *Nature Med.* **8**, 831–840 (2002).
- Meerpohl, H. G., Lohmann-Matthes, M. L. & Fischer, H. Studies on the activation of mouse bone marrow-derived macrophages by the macrophage cytotoxicity factor (MCF). *Eur. J. Immunol.* **6**, 213–217 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from FWO (G.0726.10), Belgium, and from VIB. The authors are thankful to P. Carmeliet for the *Phd2* KO and cKO mice, M. Karin for the *Ikbkb* cKO mice, P. Ratcliffe for the PHD2^{H313A} construct, A. Luttun and P. Fazzari for comments, Y. Jonsson and T. Janssens for technical assistance. VE-Cadherin:CreERT and PDGFRB:Cre transgenic mice were generated at Cancer Research UK and kindly donated by R. Adams. E.D. was funded by ARC, S.C. by FCT, R.L.O. and Y.F. by FWO, A.H. by DFG. C.R. was supported by COST action TD0901. M.D.P. was supported by an ERC starting grant.

Author Contributions Y.T., E.D. and S.C. performed experimental design, all experiments, acquisition of data and analysis and interpretation of all data. C.R. performed analysis of histological stainings, angiographies. R.L.O., C.R. and S.C. performed the western blots. R.L.O. and V.F. performed treadmill-running tests, quantitative PCR experiments and drug administrations. M.L.S. performed lentiviral vector preparation and cell transduction. F.B. performed EC isolation and angiography measurements. J.M., B.F.J. and B.G. performed oxymetry experiments. S.D. performed luciferase assays. M.W. and A.H. performed transplantation experiments and electroporations. Y.T. and J.S. performed the ligations of the femoral artery. Z.W.Z. and M.S. performed micro-computer tomography angiograms. A.A. and K.A. contributed vital reagents. T.B. and P.M. contributed in generating the *Phd2* targeting vector. Y.T., E.D., S.C., C.R., Y.F. and M.D.P. participated in scientific discussion and drafting of the manuscript. M.M. performed experimental design, data analysis, conducted scientific direction and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.M. (massimiliano.mazzone@vib-kuleuven.be)

METHODS

Animals. 129/S6 or Balb/c, female and male, WT and *Phd2*^{+/-} mice (8–12 weeks old) were obtained from our mouse facility. *Phd2*^{+/-} and *Phd2* conditional knockout mice were obtained as previously described¹. VE-Cadherin:CreERT and PDGFRB:Cre transgenic mice were generated by R. Adams at the Cancer Research UK^{29,30}. *Ikbkb* conditional knockout mice were obtained from M. Karin³¹. Tie2:Cre and Rosa26:CreERT transgenic mice were purchased from the Jackson Laboratory. Housing and all experimental animal procedures were performed in accordance with Belgian law on animal care and were approved by the Institutional Animal Care and Research Advisory Committee of the K. U. Leuven (P036/2009).

Mouse model of hindlimb ischaemia and myocardial infarction. To induce hindlimb ischaemia, unilateral or bilateral ligations of the femoral artery and vein (proximal to the popliteal artery) and the cutaneous vessels branching from the caudal femoral artery side branch were performed without damaging the nervus femoralis. By using this procedure, collateral flow to adductor muscles is preserved via arterioles branching from the femoral artery, therefore 50% up to 60% of perfusion is preserved by this method. Two superficial pre-existing collateral arterioles, connecting the femoral and saphenous artery, were used for analysis. Functional perfusion measurements of the collateral region were performed using a Liscia PIM II camera (Gambro). Gelatin-bismuth-based angiography was performed as described before and analysed by photoangiographs (Nikon D1 digital camera). Collateral side branches were categorized as follows: secondary collateral arterioles directly branched from the primary collateral, and tertiary arterioles orientated perpendicularly to the secondary branches. The number of secondary and tertiary collateral arterioles was counted. After perfusion-fixation, the muscle tissue between the two superficial collateral arterioles (adductor) was post-fixed in 2% paraformaldehyde, paraffin-embedded and morphometrically analysed²⁷. An endurance treadmill-running test was performed at baseline and 12 h post-bilateral-ligation. Myocardial infarction was induced by permanent ligation of the left anterior descending coronary artery as previously described³². Briefly, the left thorax of anesthetized mice was opened in the fourth intercostal space and all the muscles overlying the intercostal region were dissected. The main left anterior descending coronary artery was ligated proximal to the main bifurcation through a small incision of the pericardium. Discoloration of the ventricle after blood flow restriction was used as readout of a successful surgical procedure. Gelatin-bismuth-based angiography was performed 24 h post-ligation and the entire heart was fixed in 2% paraformaldehyde.

Oxymetry. Oxygen tension (*pO*₂) in the lower limb was measured using ¹⁹F-MRI oxymetry in non-ligated and ligated legs 12 h after femoral artery ligation. The oxygen reporter probe hexafluorobenzene was injected directly into the crural muscle. Magnetic resonance imaging was performed with a 4.7 T (200 MHz, ¹H), 40-cm inner diameter bore system (Bruker Biospec). A tunable ¹H/¹⁹F surface coil was used for radiofrequency transmission and reception³³.

Histology, immunostaining, and morphometry. Adductor, crural muscles and hearts were dissected, fixed in 2% paraformaldehyde, dehydrated, embedded in paraffin, and sectioned at 7-μm thickness. Area of necrotic tissues in the crural muscle was analysed by haematoxylin and eosin staining. Necrotic cells display a more glassy homogeneous appearance in the cytoplasm with increased eosinophilia, whereas the nuclear changes are reflected by karyolysis, pyknosis and karyorrhexis. Necrotic area was defined as the percentage of area which includes these necrotic myocytes, inflammatory cells and interstitial cells, compared to the total soleus area. Infarct size was measured in desmin-stained hearts 24 h after ischaemia as previously described³⁴. After deparaffinization and rehydration, sections were blocked and incubated overnight with primary antibodies: rat anti-CD31, dilution 1/500 (BD-pharmingen); mouse anti-αSMA, dilution 1/500 (Dako); rat anti-F4/80, dilution 1/100 (Serotec); dilution 1/50 (BD-pharmingen); rat anti-CD45, dilution 1/100 (BD-pharmingen); goat anti-MRC1, dilution 1/200 (R&D); rabbit anti-desmin dilution 1/150 (Cappel). To analyse capillary density and area, images of anti-CD31-stained sections of the entire soleus were taken at ×40. To measure bismuth-positive vessel density and area, haematoxylin- and eosin-stained paraffin sections were analysed and vessels filled with bismuth-gelatin (black spots) were taken in account. Images of the entire soleus were acquired at ×20 for this analysis. The values in the graph represent the averages of the mean vessel density and area per soleus muscle. The same method was used to quantify vessel capillaries and collateral branches in cardiac tissues. Density and area were measured by using a KS300 (Leica) software analysis. Hypoxic cells were analysed 2 h after injection of 60 mg kg⁻¹ pimonidazole into operated mice. Mice were killed and muscles were collected. Paraffin sections were stained with Hypoxiprobe-1-Mab-1 (Hypoxiprobe kit; Chemicon International) following the manufacturer's instructions. Oxidative stress and proliferation rate were assessed on 7-μm thick cryosections by using the goat anti-8-OHdG antibody, dilution 1/200 (Serotec) and the rat anti-BrdU antibody, dilution 1/300 (Serotec). Sections were subsequently incubated with appropriate secondary

antibodies, developed with fluorescent dyes or 3,3'-disminobenzidine (DAB, Sigma). Whole-muscle viability was assessed on unfixed 2-mm thick tissue slices by staining with 2,3,5-triphenyltetrazolium chloride (TTC). Viable area, stained in red, was traced and analysed. Pictures for morphometric analysis were taken using a Retiga EXi camera (Q Imaging) connected to a Nikon E800 microscope or a Zeiss Axio Imager connected to an Axiocam MRC5 camera (Zeiss) and analysis was performed using KS300 (Leica). Angiograms were obtained by X-ray and micro-computed tomography angiographies of hearts and legs at baseline.

Macrophage preparation. To harvest peritoneal macrophages (pMØ), the peritoneal cavity was washed with 5 ml of RPMI 10% FBS. The pooled cells were then seeded in RPMI 10% FBS in 6-well plates (2 × 10⁶ cells per well), 12-well plates (1 × 10⁶ cells per well), or 24-well plates (5 × 10⁵ cells per well). After 6 h of incubation at 37 °C in a moist atmosphere of 5% CO₂ and 95% air, non-adhering cells on each plate were removed by rinsing with phosphate-buffered saline (PBS). The attached macrophages were grown in different media for 12 h or 48 h depending on the experiments performed, as described below. When high amounts of cells were needed (analysis for HIF accumulation and NF-κB activity), macrophages were derived from bone marrow (BM) precursors as described before²⁸. Briefly, BM cells (2 × 10⁶ cells per ml) were grown in a volume of 5 ml in a 10-cm Petri dish (non tissue culture treated, bacterial grade) for 7 to 10 days in DMEM supplemented with 20% FBS and 30% L929 conditioned medium as a source of M-CSF. The cells obtained in those cultures are uniformly macrophages.

Quantitative PCR analysis. In order to investigate gene expression in pMØ, quantitative real-time PCR (qPCR) was performed. After preparing pMØ, the cells were cultured in normoxic condition for 12 h and RNA was extracted. To analyse the expression levels of chemoattractants in the adductor, tissues were collected at baseline or 24 h post-ischaemia and RNA was extracted. Macrophages and endothelial cells were freshly sorted from dissected adductors as described below and RNA was extracted. Quantitative PCR was performed with commercially available or home-made primers and probes for the studied genes. The assay ID (Applied Biosystems) or the sequence of primers and probes (when custom-made) are listed in Supplementary Table 4. RNA levels of *Sdf1* and *Pdgfb* after inhibition of NF-κB pathway were measured by qPCR on pMØ exposed for 12 h to 500 nM 6-amino-4-(4-phenoxyphenylethylamino)quinazoline.

Protein extraction and immunoblot. Protein extraction was performed using 8 M urea buffer (10% glycerol, 1% SDS, 5 mM DTT, 10 mM Tris-HCl, pH 6.8) as previously described¹. Nuclear proteins were extracted in 1% SDS buffer upon cytoplasmic separation by using a hypotonic lysis buffer (10 mM KCl, 10 mM EDTA, 0.5% NP40, 10 mM HEPES, pH 8, supplemented with phosphatase and protease inhibitors, from Roche). Signal was detected using the ECL system (Invitrogen) according to the manufacturer's instructions. The following antibodies were used: rabbit anti-HIF-1α (Novus), rabbit anti-p105/50, rabbit anti-HIF-2α (Abcam), PM9 rabbit anti-HIF-2α (from P.M.), mouse anti-vinculin (Sigma), rabbit anti-lamin A/C, rabbit anti-p65 (Cell Signaling). Densitometric analysis was performed using ImageJ 1.44 (<http://rsbweb.nih.gov/ij/>).

Transduction and transfection of BM-derived macrophages and lung endothelial cells. To express an inducible NF-κB-responsive firefly luciferase reporter, commercially available lentiviral vectors (LV) were used (Cignal Lenti NF-κB Reporter; SABiosciences). BM-derived macrophages (2.5 × 10⁵) and 10⁵ primary lung endothelial cells, isolated as described before¹, were seeded in a 24-well plate in DMEM 10% FBS or M199 20% FBS for 8 h. Cells were transduced with 10⁸ transducing units per ml. Eight hours after transduction, the medium was replaced. After 48 h, cells were stimulated with TNF-α (20 ng ml⁻¹) for 8 h and the same amount of protein extract was read in a luminometer. For PHD3 silencing, siRNA oligonucleotides were designed using the Invitrogen online siRNA design tool (<http://rnaidesigner.invitrogen.com>). The following siRNA sequences (sense strands) were used. For *Egln3/Phd3* (NM_028133.2): 5'-GCCGGCTGGGCAAAATCTATGTCA-3'; for the scramble: 5'-CACCGCTTAACCCGATTGTCCTAT-3'. In brief, one day after the transduction of macrophages with LV, cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Preparation of the oligonucleotide-Lipofectamine 2000 complexes was done as follows: 25 pmol siRNA oligonucleotide (stock 20 μM) was diluted in 50 μl Opti-MEM I reduced serum medium. Lipofectamine 2000 (1.5 μl) was diluted in 50 μl Opti-MEM I reduced serum medium and incubated for 5 min at room temperature. siRNA oligonucleotides were gently mixed with Lipofectamine 2000 and allowed to incubate at room temperature for 20 min to form complexes. Just before transfection, the cell culture medium was removed and cells were rinsed twice with serum-free Opti-MEM I medium. The Lipofectamine 2000-siRNA oligonucleotide complexes were added to each well in 400 μl of serum-free Opti-MEM medium for 5 h. Afterwards, cells were incubated in complete medium for 48 h at 37 °C in a CO₂ incubator and assayed for gene knockdown (qPCR) and luciferase activity. To assess if the increased NF-κB activity observed in *Phd2*^{+/-} macrophages was dependent on the hydroxylase activity of PHD2, 48 h before transduction,

4×10^6 BM-derived macrophages were resuspended in 240 μ l of Opti-MEM and were electroporated (250 V, 950 μ F, ∞ Ω) with 7 μ g of plasmids expressing a wild-type PHD2 (PHD2^{WT}) or a PHD2 containing a mutation at the catalytic site (PHD2^{H313A}). Silencing of the canonical pathway subunits p65 (*Rela*) and p50 (*Nfkb1*) was achieved by electroporation with specific siRNAs. Briefly, 48 h before transduction, 2.4×10^6 BM-derived macrophages were resuspended in 320 μ l of Opti-MEM and were electroporated (250 V, 950 μ F, ∞ Ω) with 60 pmol of siRNA for either scramble, p65, p50, or combination of p50 and p65. For higher efficiency of silencing, two different siRNA sequences for each respective gene were designed (<http://rnaidesigner.invitrogen.com>). For p65 (NM_009045.4): 5'-TGTCTGCACC TGTTCACAAATT-3' and 5'-TGCTGATGGAGTACCCTGATT-3'; for p50 (NM_008689.2): 5'-GAATACITCATGTGACTAATT-3' and 5'-CAAAGGTTAT CGTTCAGTTTT-3'; for the scramble: 5'-CACCGCTTAACCCGATTGCCTAT-3'.

Cell migration and viability assays. Migration and proliferation of smooth muscle cells (SMCs) and endothelial cells were assessed by using 8- μ m-pore Transwell permeable plate for migration assays and 0.4- μ m-pore Transwell permeable plate for proliferation assays (Corning Life Science). To determine cell migration and proliferation in response to soluble factors secreted by pMØ, pMØ were pre-cultured in the lower chamber for 12 h in RPMI 1% FBS or in M-199 1% FBS (migration assay), or 48 h in DMEM-F12 1% FBS or in M-199 1% FBS (proliferation assay). For migration assays, hCASMCs (human coronary artery SMCs; from Lonza) and HUVECs (human umbilical vein endothelial cells; from Lonza) were starved for 12 h in their own medium at 1% FBS and then seeded in the upper chamber (5×10^3 cells in 200 μ l of medium 1% FBS). SMCs and HUVECs were incubated for 2 days or 24 h, respectively, and migrated cells were fixed with 4% paraformaldehyde, stained with 0.25% crystal violet/50% methanol and counted under the microscope. For cell growth assays, RAOSMCs (rat aortic SMCs) and HUVECs were seeded on the upper chambers (5×10^3 cells per transwell) and cultivated with pMØ for 24 h in DMEM-F12 1% FBS or M-199 1% FBS for RAOSMCs and HUVECs, respectively. The cell proliferative ability was then analysed using WST-1 Cell Proliferation Assay (Roche Applied Biosciences) according to the manufacturer's instructions. Alternatively, WT and *Phd2*^{+/-} pMØ were seeded in the lower chamber of a Transwell and transduced with lentiviral vectors (10^8 TU/ml; Sigma) carrying an shRNA against *Sdf1* (NM_013655.4), *Pdgfrb* (NM_011057.3), or a scramble control (purchased from Sigma; TRCN0000178772 for *Sdf1*: 5'-CCGGCTGAAGAACAACAACAGACAA CTCGAGTTGTCTGTTGTTCTTTCAGTTTTTG-3'; TRCN0000042529 for *Pdgfrb*: 5'-CCGGGAGTCGAGTTGGAAGCTCATCTCGAGATGAGCT TTCCAACCTCGACTCTTTTG-3'; SHC002V for the scramble: 5'-CCGGCA ACAAGATGAAGAGACCAACTCGAGTTGGTGCTCTTCATCTTGTGTT TTT-3'). Sixty hours after macrophage transduction, SMC migration or growth assays were performed by seeding the SMCs in the upper side of the Transwell as above.

SMC differentiation assay. pMØ were seeded in a 24-well plate with DMEM F-12 5% FBS. Conditioned medium was collected after 2 days and supplemented with

25 mM HEPES. RAOSMCs were seeded in a 24-well plate (80×10^3 cells per well) and incubated for 5 h at 37 °C in a moist atmosphere of 5% CO₂ and 95% air. After 2 h of starvation in DMEM-F12 1% FBS, SMCs were stimulated with conditioned medium from WT and *Phd2*^{+/-} pMØ. After 24 h, differentiation status of the SMCs was assessed by qPCR.

Macrophage and endothelial cell sorting. For cell sorting of adductor macrophages and endothelial cells, the adductors were dissected, dissociated mechanically, and digested using collagenase I for 45 min at 37 °C. For macrophage sorting, the digested cell suspension was incubated for 15 min with mouse anti-CD16/CD32 mAb (Fc Block, BD-pharmingen) and stained with rat FITC-conjugated anti-F4/80 antibody (Serotec) for 20 min at 4 °C. CD31⁺CD45⁻ endothelial cells were sorted from the digested adductor cell suspension after incubation with rat APC-conjugated anti-CD31 and rat FITC-conjugated anti-CD45 (BD-pharmingen) for 20 min at 4 °C.

BM transplantation and haematological analysis. Balb/c WT and *Phd2*^{+/-} recipient mice were irradiated with 7.5 Gy. Subsequently, 5×10^6 BM cells from green fluorescent protein⁺ (GFP⁺) WT or GFP⁺ *Phd2*^{+/-} mice were injected intravenously via the tail vein. Femoral artery ligation, treadmill running test and bismuth angiography were performed 5 weeks after BM reconstitution. Red and white blood cell count was determined using a haemocytometer (Cell-Dyn 3700, Abbott) on peripheral blood collected in heparin by retro-orbital bleeding. To assess the effect on arteriogenesis of acute deletion of one *Phd2* allele in macrophages, 5×10^6 BM cells from *Phd2*^{Rosa26CreERT²lox/WT} mice were transplanted into lethally irradiated WT recipient mice. After 5 weeks, transplanted mice were injected intraperitoneally with tamoxifen (1 mg per mouse; Sigma) or vehicle for 5 days. Femoral artery ligation was performed 10 days after tamoxifen treatment as above.

Statistics. The data were represented as mean \pm s.e.m. of the indicated number of measurements. Statistical significance was calculated by two-tailed unpaired *t*-test for two data sets and ANOVA followed by Bonferroni post hoc test for multiple data sets using Prism (GraphPad Inc.), with *P* < 0.05 considered statistically significant.

29. Foo, S. S. *et al.* Ephrin-B2 controls cell motility and adhesion during blood-vessel-wall assembly. *Cell* **124**, 161–173 (2006).
30. Benedito, R. *et al.* The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137**, 1124–1135 (2009).
31. Chen, L. W. *et al.* The two faces of IKK and NF- κ B inhibition: prevention of systemic inflammation but increased local injury following intestinal ischemia-reperfusion. *Nature Med.* **9**, 575–581 (2003).
32. Lutgens, E. *et al.* Chronic myocardial infarction in the mouse: cardiac structural and functional changes. *Cardiovasc. Res.* **41**, 586–593 (1999).
33. Jordan, B. F., Cron, G. O. & Gallez, B. Rapid monitoring of oxygenation by 19F magnetic resonance imaging: Simultaneous comparison with fluorescence quenching. *Magn. Reson. Med.* **61**, 634–638 (2009).
34. Pfeffer, M. A. *et al.* Myocardial infarct size and ventricular function in rats. *Circ. Res.* **44**, 503–512 (1979).

Perception of sniff phase in mouse olfaction

Matthew Smear^{1,2}, Roman Shusterman¹, Rodney O'Connor^{1†}, Thomas Bozza^{1,2} & Dmitry Rinberg¹

Olfactory systems encode odours by which neurons respond and by when they respond^{1–3}. In mammals, every sniff evokes a precise, odour-specific sequence of activity across olfactory neurons^{4–6}. Likewise, in a variety of neural systems, ranging from sensory periphery^{7,8} to cognitive centres⁹, neuronal activity is timed relative to sampling behaviour and/or internally generated oscillations. As in these neural systems, relative timing of activity may represent information in the olfactory system^{10,11}. However, there is no evidence that mammalian olfactory systems read such cues^{12,13}. To test whether mice perceive the timing of olfactory activation relative to the sniff cycle ('sniff phase'), we used optogenetics in gene-targeted mice to generate spatially constant, temporally controllable olfactory input. Here we show that mice can behaviourally report the sniff phase of optogenetically driven activation of olfactory sensory neurons. Furthermore, mice can discriminate between light-evoked inputs that are shifted in the sniff cycle by as little as 10 milliseconds, which is similar to the temporal precision of olfactory bulb odour responses^{14,15}. Electrophysiological recordings in the olfactory bulb of awake mice show that individual cells encode the timing of photoactivation in relation to the sniff in both the timing and the amplitude of their responses. Our work provides evidence that the mammalian olfactory system can read temporal patterns, and suggests that timing of activity relative to sampling behaviour is a potent cue that may enable accurate olfactory percepts to form quickly^{11,16}.

If mice perceive the timing of olfactory activation, they should be able to discriminate between identical sensory stimuli presented at different times in the sniff cycle. In order to isolate this cue, we used optogenetics¹⁷ to deliver spatially fixed, temporally controllable patterns of olfactory sensory neuron (OSN) stimulation. We engineered a mouse line in which all OSNs express channelrhodopsin-2 fused to the yellow fluorescent protein (ChR2–YFP) from the Olfactory Marker Protein (OMP) locus (Fig. 1a). In OMP–ChR2 mice, ChR2–YFP is expressed in all mature olfactory sensory neurons and their nerve terminals in glomeruli of the olfactory bulb (Fig. 1b).

To establish that we can stimulate the olfactory system with light in these mice, we first tested light detection in OMP–ChR2 ($n = 12$) and wild-type littermate controls ($n = 4$). We implanted a pressure cannula into one nasal cavity to measure sniffing, and an optical fibre in the contralateral cavity for photostimulation (Fig. 1c; see Methods). We tested these mice in a head-fixed, go/no-go task in which mice report perceptual judgments by licking or not (Fig. 1d; see Methods). We first trained mice to report odour detection. All mice achieved above-chance behavioural performance in their first session (binomial test, $P < 0.01$, 200–400 trials), and performed $>90\%$ in subsequent sessions (Fig. 1e, Supplementary Fig. 1). After at least four odour sessions, we replaced odour stimuli with light pulses (5 mW power, 1 ms duration). Under these conditions, all OMP–ChR2 mice reported detection of light with similar accuracy as for odour, within the first session (Fig. 1e), while all wild-type mice failed to report light detection above chance level in any of four sessions (binomial test, $P > 0.05$). This shows that light drives behaviour through ChR2-mediated OSN activation.

To test whether the animals perceive the sniff phase of OSN activation, we trained mice ($n = 8$) to discriminate between light stimuli solely on the basis of this cue. In each trial, a single light stimulus occurred, and across trials, stimulus intensity and duration were held constant. Stimuli were delivered 32 ms after the onset of inhalation ('go' sniff phase) or 32 ms after the onset of exhalation ('no-go' sniff phase; Fig. 2a, Supplementary Fig. 2a; see Methods). After switching

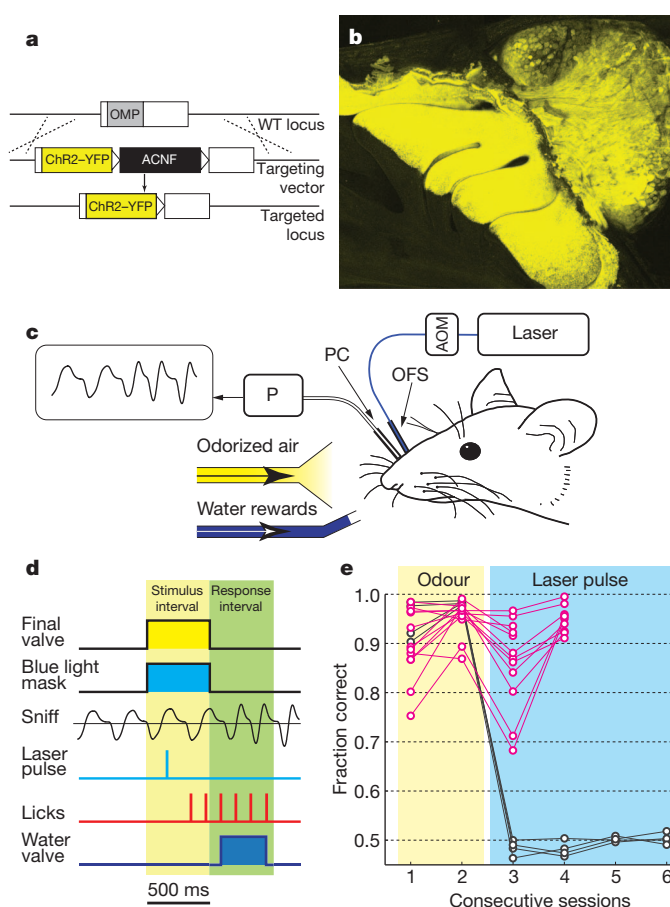


Figure 1 | Stimulating olfaction with light. **a**, Diagram of the gene targeting strategy. The ChR2–YFP sequence (yellow box) replaces that of OMP (grey box). The targeting selection cassette (ACNF) was removed by germline excision, leaving behind a single loxP site (triangle). **b**, Sagittal view of whole-mount olfactory epithelium and bulb. In OMP–ChR2 mice, ChR2–YFP labels OSNs and their axons in the bulb. **c**, Schematic of experimental set-up. Mice were implanted with a nasal optical fibre stub (OFS) to deliver light, gated by an acousto-optic modulator (AOM). A nasal pressure cannula (PC) coupled to a pressure sensor (P) measures sniffing. Inverted intranasal pressure signal is shown at top left. **d**, Behavioural trial structure. Each trial comprises a stimulus interval (yellow shading) and a response interval (green). **e**, Performance of OMP–ChR2/+ mice (pink circles; $n = 12$) and +/+ littermate controls (black circles; $n = 4$) in odour detection sessions (yellow shading), followed by light detection sessions (blue shading).

¹Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. ²Department of Neurobiology, Northwestern University, Evanston, Illinois 60208, USA. †Present address: Department of Biology, Boston University, Boston, Massachusetts 02215, USA.

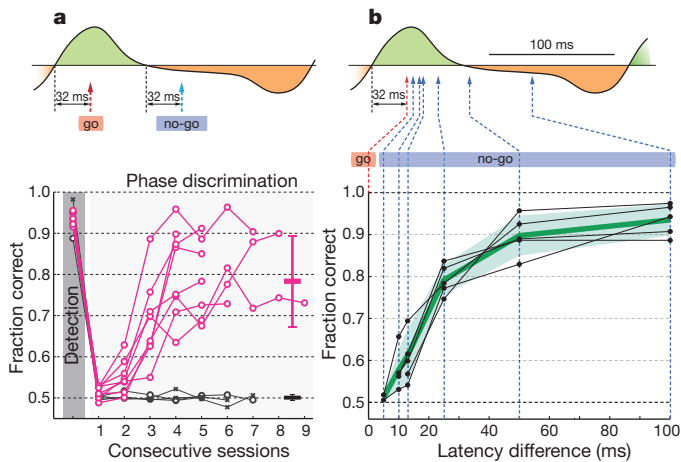


Figure 2 | OMP-ChR2 mice perceive sniff phase. **a**, Top, schematic of the sniff phase discrimination problem, shown relative to a typical sniff waveform, with inhalation shaded green and exhalation shaded orange. Light was delivered 32 ms after inhalation onset (red arrow) in 'go' trials or 32 ms after exhalation onset (blue arrow) in 'no-go' trials. Bottom, connected pink circles show the performance of OMP-ChR2/+ mice ($n = 8$) for their last light detection session, followed by phase discrimination sessions. Black lines show click detection (grey shading) and click phase discrimination performance for OMP-ChR2/+ (black circles, $n = 2$) and wild-type (black asterisks, $n = 2$). Horizontal dashes at right show mean \pm s.d. phase discrimination performance for light (pink; $n = 30$ sessions) and click (black; $n = 19$ sessions) stimuli, from the third session onward. **b**, Performance as a function of latency difference. Top, as **a**. Bottom, black filled circles show performance of individual OMP-ChR2 mice while green line and shaded region give mean \pm s.d. ($n = 5$).

from detection to the phase discrimination task, mice learned quickly, attaining above-chance performance in their second or third session (Fig. 2a). This behavioural performance demonstrates that mice perceive the sniff phase of olfactory input. In contrast, another set of mice ($n = 4$), tested with an easily detectable auditory click stimulus, failed to report the sniff phase of clicks (Fig. 2a, binomial test, $P > 0.05$). This suggests that the olfactory system may have unique access to sniff timing information.

How acute is the mouse's sense of time in the sniff? To test whether mice can discern finer timing differences, we trained five mice to discriminate between light stimuli occurring at the same 'go' sniff phase as above, and those occurring with varying latencies (5–100 ms) after the 'go' sniff phase (Fig. 2b, top). A single 'no-go' latency was tested in each session. Performance is high for 'no-go' latencies of 50 ms or greater (Fig. 2b). Mice maintained high accuracy at 25 ms latency (Supplementary Fig. 2b; $80 \pm 5\%$, mean \pm s.d.), and four of five mice exceeded chance performance at 10 ms (binomial test, $P < 0.01$). Achieving this performance does not require extensive training—at each latency, all mice performed three or fewer sessions. Across sessions, sniff durations do not differ systematically, but do vary from trial to trial, and mice performed better in trials with short inhalation duration (Supplementary Fig. 3). These data show that mice can sense timing differences that are tenfold shorter than a sniff cycle.

To characterize how the olfactory bulb responds to the stimuli presented in our behavioural experiments, we recorded light-evoked responses from 86 neurons, putatively mitral/tufted (M/T) cells, in the olfactory bulb of five OMP-ChR2 mice (see Methods). These mice were awake but were not performing a task. Out of 86 cells, 57% exhibited light-evoked responses: 26 gave excitatory responses, while 23 cells gave inhibitory responses (see Methods). By comparison, in a recent study, it was found that individual odours, on average, evoke responses in 66% of M/T cells¹⁵. Therefore, the light stimulus used in our behavioural experiments activates a similar number of M/T cells as do odours.

We then delivered light stimuli at a range of latencies relative to sniff (2–6 latencies per recording session). Some cells responded strongly at

all latencies tested (for example, Fig. 3a, cell 1). In contrast, other cells exhibited varying response amplitude with stimulus latency (for example, Fig. 3a, cell 2).

To quantify the temporal dynamics of excitatory light responses, we fitted a Gaussian function to the difference between inhalation-aligned spike histograms with and without stimuli. The fit parameters yield measures of latency (τ), duration (σ) and amplitude (A) (see Methods). The brief durations (Fig. 3b) and narrow latency distribution (Fig. 3c) of these responses demonstrate that M/T cells faithfully propagate the timing of OSN photostimulation to their central targets. In addition, cells may vary their response amplitudes when stimulated at different times in the sniff cycle (for example, Fig. 3a, cell 2). Tuning curves for latency relative to sniff were heterogeneous across cells and often non-monotonic (Fig. 3c; tuning curves for cells receiving stimuli at six latencies are shown in Supplementary Fig. 4). As a result of this tuning, information about timing of OSN activation is also contained in the pattern of response amplitudes across the M/T cell ensemble. Consequently, olfactory bulb responses contain two cues that may enable the animal to report the latency of light stimuli relative to sniff onset: timing and amplitude.

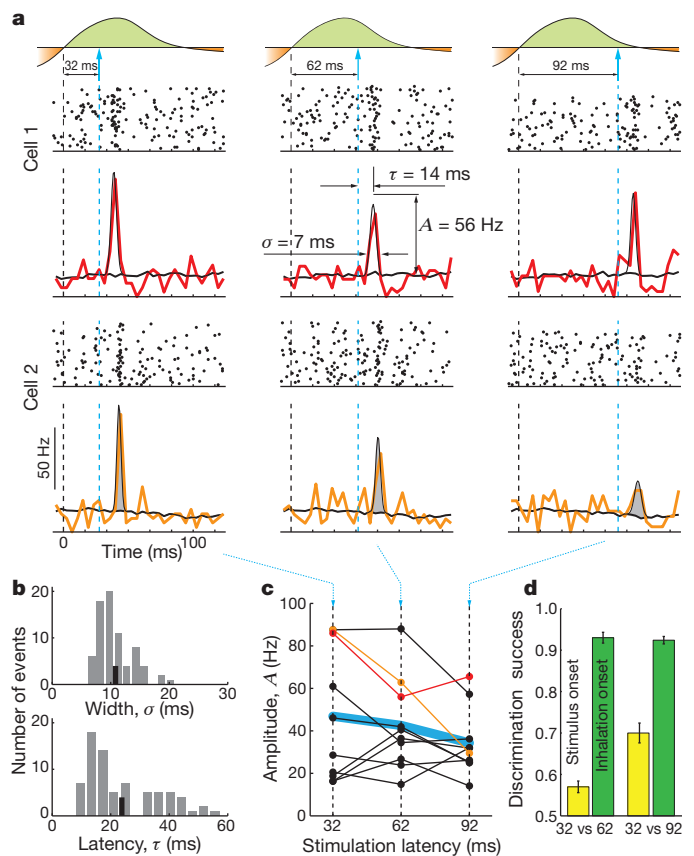


Figure 3 | Response of mitral/tufted cells to light stimulation. **a**, Top, light application. Middle and bottom, raster plots + PSTH for two cells' responses (cell 1, top; cell 2, bottom) to light at three latencies (32 ms, left; 62 ms middle; 92 ms, right) after inhalation onset. Coloured lines are PSTHs for light responses. Thick grey lines are PSTHs for spontaneous activity. Thin black line is a Gaussian fit of the difference between PSTHs for stimulated and unstimulated sniffs. The fit parameters yield measures of response width (σ), latency (τ) and amplitude (A). **b**, Distribution of response widths (σ , top) and latencies (τ , bottom). Grey bars, data; thick black line, mean. **c**, Connected filled circles show response amplitudes (A) from individual cells (red and orange dots show respectively cells 1 and 2). Blue line indicates the across-cell mean. **d**, Classification performance for the neuronal population response (mean \pm s.d. across repeated permutations) discriminating between 32 and 62 ms and between 32 and 92 ms light stimulation latency. Responses were aligned to the stimulus onset (yellow) and inhalation onset (green).

We estimated the amount of information that neuronal response timing and amplitude carry about stimulus sniff-timing using a classification algorithm. In essence, this algorithm measures how well stimulus sniff phases can be discriminated on the basis of each trial's neuronal response cues (51 cells; see Methods). We compared classification success with and without the temporal cue, by applying the classification algorithm to the same neural data aligned in two ways: to inhalation onset (as in Fig. 3a), in which case both the sniff-timing and amplitude cues are available, or to stimulus onset, which eliminates the timing cue. When the classification algorithm is applied to sniff-aligned data, discrimination success for both pairs of stimuli is above 90% (Fig. 3d), comparable to behavioural performance (Fig. 2b). When applied to stimulus-aligned data, the classification algorithm performs worse, yet still above chance for both pairs of stimuli (Fig. 3d; 32 versus 62: $57 \pm 1.4\%$; 32 versus 92: $70 \pm 2.4\%$; see Methods). These analyses indicate that both timing and amplitude cues carry information about stimulus latency relative to sniff.

Our work provides evidence (the first, to our knowledge) that the mammalian olfactory system can read temporal information and reveals the striking temporal acuity of this mechanism. Mice discriminate between inputs solely on the basis of 'sniff phase'—that is, relative time in the sniff cycle. Whether this temporal cue is read in phase or time coordinates remains an open question^{14,15}. Although mice can use timing relative to sniff in our experimental design, where this is the only cue available, it is unclear to what degree animals do so in the context of natural odour stimulation. However, the ease with which mice sense the sniff phase of olfactory input in isolation argues that this cue plays an important role in representing odours.

What mechanisms temporally pattern olfactory responses? Recent work in insects has revealed odour-specific temporal structure in OSN responses^{18,19}, which can largely be explained by a simple kinetic model of ligand-receptor binding¹⁹. Rodent OSNs also give temporally structured responses: calcium-imaging studies demonstrate odour-specific sniff-locked sequences across glomeruli⁶. These sniff-locked sequences may propagate to recipient neurons in the olfactory bulb, where additional sniff-entrained circuit mechanisms may operate. For example, M/T cells exhibit sniff-locked oscillations of membrane potential^{16,20}, which can transform input intensity into timing by modulating excitability^{10,21}. Consistent with this idea, raising odour concentration shifts activation to earlier times in the sniff cycle^{16,20}. The resultant mapping of intensity to time agrees with a theory in which latency relative to the sniff onset encodes the intensity of receptor activation^{10,16,22}. Determining whether these or other mechanisms account for how the olfactory system transforms odours to sniff-locked temporal patterns will require further investigation. We provide a strong impetus for such work by showing that mice can perceive sniff phase cues.

The behavioural relevance of temporal coding has received experimental support in a variety of sensory systems^{23–25}. Despite the long history of work on the temporal patterning of odour responses^{4–6}, the behavioural relevance of temporal coding in olfaction has not been demonstrated, despite a previous attempt to do so¹². Our strategy of timing optogenetic activation relative to a putative timing reference signal, which has proven successful for olfaction, could be generalized to test phase/latency coding hypotheses in other systems²⁶. For example, timing relative to sampling behaviour and/or local field potential oscillations has been proposed as a coding variable in vibrissa somatosensation²⁷ and vision^{8,28}. Beyond more peripheral sensory areas, our optogenetic strategy may also be applicable to putative temporal cues in more central systems—for example, the phase precession of hippocampal place cells⁹.

The ability of mice to discriminate between identical patterns of illuminated OSNs solely on the basis of timing with respect to the sniff suggests that differences in the spatial pattern of OSN activation may be unnecessary for perceptible olfactory differences, contrary to prior suggestions¹³. However, stimulus- and sniff-related signals may converge within OSNs, creating sniff-phase-dependent spatial patterns of

OSN activation. In some OSNs, activity is modulated by the sniff cycle in the absence of overt odour stimulation, perhaps responding to air-flow or background odour²⁹. Downstream of OSNs, stimulus- and sniff-related signals may be integrated by olfactory bulb circuits, or further downstream in the olfactory system. These considerations underscore the fact that spatial coding and temporal coding are not mutually exclusive, and may instead exhibit synergy in numerous ways. We speculate that time comparisons across glomeruli give a concentration-invariant readout for odour identity^{11,16,22}, whereas temporal comparison to an internal representation of the sniff yields information about odour concentration. Such a coding scheme can rapidly resolve ambiguities that arise as odour identity and intensity change¹¹. Extracting both parameters on a sniff-by-sniff basis may help animals locate and identify odour sources in natural olfactory scenes.

METHODS SUMMARY

OMP-ChR2 heterozygous mice and wild-type mice were implanted with headbars for head fixation, pressure cannulae in the nasal cavity for sniff recording and optic fibre stubs in the contralateral nasal cavity for light stimulus delivery. After at least three days of recovery followed by at least ten days of restriction of water intake to 1 ml d^{-1} , mice were trained to lick for water while head-fixed in a behavioural box. Then, mice were trained to perform the following behavioural tasks in a go/no-go paradigm: (1) odour detection; lick in response to odour and do not lick to blank delivery, (2) light detection; lick in response to light stimulation via nasal optic fibre (5 mW power, 1 ms duration), (3) sniff phase discrimination; lick in response to light stimulation at fixed latency after inhalation onset (32 ms) ('go' stimulus) and do not lick in response to light stimulation triggered after exhalation onset ('no-go' stimulus); and (4) fine temporal discrimination: lick in response to light stimulation at fixed latency after inhalation onset ('go' stimulus), and do not lick in response to light stimulation delayed from the 'go' stimulus latency by some time interval (5–100 ms, 'no-go' stimulus). In each session (~400 trials), one 'no-go' stimulus latency was used. Onsets of inhalation and exhalation were defined as zero-crossings of the intranasal pressure signal. For electrophysiological recordings, mice were also implanted with 16- or 32-channel silicon probes. In recording sessions, mice were awake but not performing a task. Light stimuli were triggered at fixed latencies (32, 62 and 92 ms) after onsets of inhalation or exhalation phase.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 April; accepted 2 September 2011.

Published online 12 October 2011.

- Laurent, G. Olfactory network dynamics and the coding of multidimensional signals. *Nature Rev. Neurosci.* **3**, 884–895 (2002).
- Friedrich, R. W. & Laurent, G. Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science* **291**, 889–894 (2001).
- Juneek, S., Kludt, E., Wolf, F. & Schild, D. Olfactory coding with patterns of response latencies. *Neuron* **67**, 872–884 (2011).
- Macrides, F. & Chorover, S. L. Olfactory bulb units: activity correlated with inhalation cycles and odor quality. *Science* **175**, 84–87 (1972).
- Chaput, M. & Holley, A. Single unit responses of olfactory bulb neurones to odour presentation in awake rabbits. *J. Physiol. (Paris)* **76**, 551–558 (1980).
- Spors, H., Wachowiak, M., Cohen, L. B. & Friedrich, R. W. Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb. *J. Neurosci.* **26**, 1247–1259 (2006).
- Szabo, T. & Hagiwara, S. A latency-change mechanism involved in sensory coding of electric fish (mormyrids). *Physiol. Behav.* **2**, 331–335 (1967).
- Gollisch, T. & Meister, M. Rapid neural coding in the retina with relative spike latencies. *Science* **319**, 1108–1111 (2008).
- O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
- Hopfield, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995).
- Schaefer, A. T. & Margrie, T. W. Spatiotemporal representations in the olfactory system. *Trends Neurosci.* **30**, 92–100 (2007).
- Monod, B., Mouly, A. M., Vigouroux, M. & Holley, A. An investigation of some temporal aspects of olfactory coding with the model of multi-site electrical stimulation of the olfactory bulb in the rat. *Behav. Brain Res.* **33**, 51–63 (1989).
- Leon, M. & Johnson, B. A. Is there a spacetime continuum in olfaction? *Cell. Mol. Life Sci.* **66**, 2135–2150 (2009).
- Cury, K. M. & Uchida, N. Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* **68**, 570–585 (2010).
- Shusterman, R., Smear, M., Koulakov, A. & Rinberg, D. Precise olfactory responses tile the sniff cycle. *Nature Neurosci.* **14**, 1039–1044 (2011).

16. Margrie, T. W. & Schaefer, A. T. Theta oscillation coupled spike latencies yield computational vigour in a mammalian sensory system. *J. Physiol. (Lond.)* **546**, 363–374 (2003).
17. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
18. Raman, B., Joseph, J., Tang, J. & Stopfer, M. Temporally diverse firing patterns in olfactory receptor neurons underlie spatiotemporal neural codes for odors. *J. Neurosci.* **30**, 1994–2006 (2010).
19. Nagel, K. I. & Wilson, R. I. Biophysical mechanisms underlying olfactory receptor neuron dynamics. *Nature Neurosci.* **14**, 208–216 (2011).
20. Cang, J. & Isaacson, J. S. In vivo whole-cell recording of odor-evoked synaptic transmission in the rat olfactory bulb. *J. Neurosci.* **23**, 4108–4116 (2003).
21. Perkel, D. H. & Bullock, T. H. Neural coding. *Neurosci. Res. Prog. Bull.* **6**, 219–349 (1968).
22. Brody, C. D. & Hopfield, J. J. Simple networks for spike-timing-based computation, with application to olfactory processing. *Neuron* **37**, 843–852 (2003).
23. Hall, C., Bell, C. & Zelick, R. Behavioral evidence of a latency code for stimulus intensity in mormyrid electric fish. *J. Comp. Physiol. A* **177**, 29–39 (1995).
24. Di Lorenzo, P. M., Leshchinskiy, S., Moroney, D. N. & Ozdoba, J. M. Making time count: functional evidence for temporal coding of taste sensation. *Behav. Neurosci.* **123**, 14–25 (2009).
25. Jacobs, A. L., Fridman, G., Douglas, R. M., Alam, N. M. & Latham, P. Ruling out and ruling in neural codes. *Proc. Natl Acad. Sci. USA* **106**, 5936–5941 (2009).
26. VanRullen, R., Guyonneau, R. & Thorpe, S. J. Spike times make sense. *Trends Neurosci.* **28**, 1–4 (2005).
27. Curtis, J. C. & Kleinfeld, D. Phase-to-rate transformations encode touch in cortical neurons of a scanning sensorimotor system. *Nature Neurosci.* **12**, 492–501 (2009).
28. Montemurro, M. A., Rasch, M. J., Murayama, Y., Logothetis, N. K. & Panzeri, S. Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Curr. Biol.* **18**, 375–380 (2008).
29. Grosmaitre, X., Santarelli, L. C., Tan, J., Luo, M. & Ma, M. Dual functions of mammalian olfactory sensory neurons as odor detectors and mechanical sensors. *Nature Neurosci.* **10**, 348–354 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Doglio and the Transgenic and Targeted Mutagenesis Laboratory at Northwestern University for generation of chimaeric mice, B. Weiland for technical help with cloning and gene targeting, D. Huber, D. O'Connor and T. Komiyama for advice on mouse behaviour, D. Wesson and M. Wachowiak for instruction on sniff measurement, J. Nunez-Iglesias for assistance with statistics, G. Shtengel for advice on laser set-up, and T. Tabachnik and H. Davidowitz for help designing the behavioural rig. J. Osborne fabricated the microdrive. G. Lott provided digital acquisition software. A. Koulakov contributed to spike-sorting and classification algorithms. We thank W. Denk, K. Svoboda, R. Gütig, R. Egnor, M. Orger and A. Resulaj for comments on the manuscript. This work was supported by the Visiting Scientist Program at JFRC. T.B. was supported by NIDCD (R01DC009640, R21DC010911), the Whitehall Foundation and the Brain Research Foundation.

Author Contributions M.S. and D.R. designed the study and build the experimental set-up, M.S. performed the experiments and analysed the behavioural data. R.S. and M.S. performed the electrophysiological recordings, R.S. and D.R. analysed the electrophysiological data, and T.B. initiated the transgenic approach and generated the gene-targeted mice. R.O. developed the laser optics and optical fibre design. M.S., T.B. and D.R. wrote the manuscript. D.R. and T.B. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.R. (rinbergd@janelia.hhmi.org) or T.B. (bozza@northwestern.edu).

METHODS

Gene targeting. The coding sequence for Chr2(H134R)-YFP (gift of G. Nagel, Max Planck Institute for Biophysics) was amplified and cloned into an OMP targeting vector³⁰ replacing the endogenous OMP coding sequence. The vector also contained an autoexcising *neo* selection cassette³¹. The vector was linearized and electroporated into E14 ES cells, and correctly targeted clones isolated using standard methods. Targeted clones were injected into C57BL/6J blastocysts to generate chimaeras. The allele was passed through the male germline, removing the *neo* cassette. The OMP-Chr2-YFP line was derived from clone 'OCY-58'. This strain will be made available through The Jackson Laboratory (Tyr<c-2J>-OMP<tm1(COP4/EYFP)-Tboz>/J; STOCK #14173); address requests for information to T.B.

Animals. Data were collected in 15 OMP-Chr2-YFP heterozygous mice and 4 wild-type littermates. All mice had at least one normal copy of OMP. Subjects were 6–8 weeks old at the beginning of behavioural training and were maintained on a 12 h light/dark cycle (lights on at 8:00 p.m.) in isolated cages in a temperature- and humidity-controlled animal facility. All animal care and experimental procedures were in strict accordance with a protocol approved by the Howard Hughes Medical Institute Institutional Animal Care and Use Committee.

Sniff recording. To monitor the sniff signal, a 7-mm-long stainless cannula (gauge 23, Small Parts capillary tubing) was implanted in the nasal cavity. The cannula was capped between experimental recordings. During experiments, the cannula was connected via polyethylene tubing (801000, A-M systems) to a pressure sensor (MPX5050, Freescale Semiconductor) and custom-made preamplifier circuit. The signal from the preamplifier was amplified 20× and low-pass-filtered below 20 Hz (Cygnus Technology), digitized with an NIDAQ board (National Instruments) and acquired by an in-house data acquisition program (SpikeHound, written by G. Lott). The pressure signal was also sent to a custom-made comparator board that created a square TTL pulse between rising and falling zero crossings in the pressure signal. This pulse went to a behavioural control board to trigger light stimuli (see below).

Surgery. Mice were anaesthetized using isoflurane gas anaesthesia. The horizontal bar for head fixation, pressure cannula, optic fibre stub, and, in a subset of mice, electrode chamber were implanted during a single session of surgery. To implant the sniffing cannula, a small hole was drilled in the bone overlying the nasal cavity, into which the cannula was inserted and affixed with glue and stabilized with dental cement. The optic fibre stub was implanted and fixed in the same way in the contralateral nasal cavity. To implant the electrode chamber, a small craniotomy (~1 mm²) was opened above the olfactory bulb, roughly centred along the A–P and M–L axes of the bulb. An electrode chamber with a silicon probe was fixed by dental cement to the skull, posterior to the olfactory bulb. The reference electrode was implanted in the cerebellum. After surgery, a mouse was caged individually and given at least 3 days for recovery.

Stimulus delivery. For odour stimulus delivery, we used a nine-odour air dilution olfactometer. Odorants (Sigma-Aldrich) were stored in liquid phase in dark vials. The airflow through the selected odourant vial was diluted 10 times by the main airflow stream and homogenized in a long thin capillary before reaching the final valve. Between stimuli, a steady stream of 1,000 ml min⁻¹ of clean air flowed to the odour port continuously, while the flow from the olfactometer was directed to an exhaust. During stimulus delivery, the final valve (four-way Teflon valve; NResearch) switched the odour flow to the odour port, and diverted the clean airflow to the exhaust. Temporal odour concentration profile was checked by a mini photoionization detector (miniPID, Aurora Scientific). The concentration reached a steady state 25–40 ms after final valve opening.

A 473-nm laser (Ciel Blue DPSS, Photonic Solutions) was our light source. The main beam was split to provide stimulus for two experiments. Each secondary beam was gated by an acousto-optic modulator (AOM, QuantaTech), which enabled analogue control of light stimulus power with microsecond timing precision. A fibre launcher (Thorlabs) was positioned to catch the first mode from the AOM in a 100 µm core multi-mode optic fibre. The amplitude of the square pulse controlled the angle of AOM beam diversion, thus providing fine control of power collected by the fibre launcher.

The opposite end of the fibre terminated in a ceramic ferrule (Precision Fibre Products), which could be coupled via an phosphor-bronze sleeve (Optequip) to an identical ceramic ferrule holding the optical fibre stub implanted into the mouse. The light stimulus power at the ferrule ending was measured by a power-meter (Thorlabs), and calibrated daily by adjusting the amplitude of the pulse to the AOM driver. The ferrule coupling allowed efficient transmission of light (80–90%), but also leaked light. To prevent the mouse from using this leaked light as a visual cue, two bright blue LEDs (Luxeon V-star, Philips Lumileds Lighting Company) were positioned on either side of the mouse's head, about 1 cm from each eye. These LEDs were activated during the stimulus period of each trial, to mask leak light from the laser.

Water delivery was based on gravitational flow controlled by a solenoid valve (Clippard) connected via Tygon tubing to a stainless steel cannula (gauge 21, Small Parts), which served as a lick tube. The lick tube was positioned near the animal's mouth, and could be moved by a micromanipulator. The water volume was controlled by the duration of valve opening for 200–400 ms duration, calibrated daily to give approximately 5 µl per opening. Licks were detected by photodiode beam break by the mouse's tongue.

Behavioural control. All behavioural events (odour and final valve opening, laser delivery, water delivery, and photobeam crossing) were monitored and controlled by a behavioural board (LASOM1, RPMetrix), which allowed real-time experimental control with millisecond precision. The behavioural board reads trial parameters and sends trial results to a PC running custom-written MatLab routines (Mathworks).

Behavioural task and training. After at least 3 days of post-operative recovery and at least 7 days of water restriction (1 ml d⁻¹), we began to train the mice. Training started with water-sampling sessions, in which the mouse was placed in the head fixation set-up and given water for licking. Before moving to the next stage of training, each mouse had to perform two sessions in which it licked enough to receive its full 1 ml of water for the day. Mice that failed to collect their full daily ration in a behavioural session were supplemented with water in their home cage.

Next, the mice were trained to report odour detection. A behavioural session was broken into pseudo-randomly ordered trials, each of which consisted of a stimulus period, a response period, and an intertrial interval (ITI). During the stimulus period, the final valve switched to direct air from the olfactometer to the animal. Olfactometer flow passed through a vial containing liquid odorant, or through a blank vial. Mice received water for licks during the response period following odour delivery, and did not receive water if they licked in response to blank delivery. These incorrect licks were punished by lengthened ITI. Correct trial ITIs were 5,000 ms plus a random number between 1 and 2,000 ms, while incorrect trial ITIs were 10,000 ms plus a random number between 1 and 6,000 ms. Randomization of ITIs was intended to prevent the possibility that animals would anticipate stimulus delivery and synchronize their sniffing.

After at least two sessions with a first odorant, another odour became the 'go' stimulus. In pilot experiments, we found that mice usually would not lick for a new stimulus under these conditions, but only for the initially trained odorant. In order to facilitate more rapid acquisition of licking for new stimuli, we included an 'associative block' at the beginning of every session. The associative block consisted of ten water valve openings delivered immediately after fixing the mouse in the behavioural chamber, followed by 20 consecutive 'go' trials with the new stimulus. Acquisition blocks were included in all behavioural sessions reported here.

Light sessions began after at least four odour detection sessions for each mouse. Licking in response to light stimulus was rewarded with water. Licking when no light stimulus was delivered lengthened the ITI. For all light sessions, the stimulus power at the ferrule coupling was 5 mW, while the stimulus duration was 1 ms. Pilot experiments suggested that stimuli roughly an order of magnitude more powerful or longer duration could be detected by +/+ mice. All OMP-Chr2/+ mice reported light detection. Only those mice that maintained a good sniff signal could be tested for the temporal discriminations.

After at least one session of light detection, mice began sniff phase discrimination sessions. In these, light was triggered from a rising or falling zero crossing in the sign-inverted pressure signal, which indicate the onset of inhalation or exhalation, respectively. Reliable detection of zero-crossing events was facilitated by low pass filtering sniff signals below 20 Hz, which introduced a constant delay of 32 ms, as described in the text. After at least three sniff phase discrimination sessions, those mice that maintained a good sniff signal were tested for the finer latency discriminations. In each of these sessions, a single no-go stimulus was used, and each mouse did three or fewer sessions at each latency, in descending order.

Electrophysiology. Mitral/tufted cell spiking activity was recorded using 16- or 32-channel silicon probes (NeuroNexus, models a2x2-tet-3mm-150-150-312(F16), a4x8-5mm 150-200-312(F32)). Cells were recorded in the mitral cell layer of the dorsal bulb, 300–400 µm from the bulb surface. The identity of M/T cells were established based on the criteria formulated in previous work³². The data were acquired using 32-channel data acquisition system (Digital Lynx, Neuralynx) with widely open broadband filters and sampling frequency 0.1–9,000 Hz.

Data analysis and spike extraction. Data analysis was done in MatLab (Mathworks). Acquired electrophysiological data were filtered and spike sorted. For Si-probe data we used the M-Clust program (written by A. D. Redish) and a software package (written by A. Koulakov).

Light responses. To identify excitatory and inhibitory responses in neurons, we used a randomization test to compare the distributions of total spike counts of each cell with and without light stimulation³³. In each session, 3,000–5,000 sniff cycles without light stimulation nor following light stimulation within 5 cycles were

defined as control cycles. N cycles with light stimulation (one sniff cycle per trial with stimulus) were test cycles ($N = 22\text{--}70$). For time windows of 1, 2, ... 100 ms after stimulus onset, we counted the number of spikes in the time window across the N trials. We then compared this number against the distribution of spike counts in the same time window for randomly chosen subsets of size N from the control cycles. The P value was estimated as the proportion of control spike counts larger than the observed test count, relative to the distribution median, multiplied by 2 to account for the two-sidedness of the test. We considered a cell to respond to the stimulus if the P value for at least one time window was less than 0.003, which corresponded to a false discovery rate of 0.05 by the Benjamini-Hochberg procedure. For statistically significant excitatory responses, we fitted a Gaussian, $f = A \exp[-\pi(t - \tau)^2/\sigma^2]$, to the difference between spike histograms for stimulated and unstimulated sniffs, where A is the amplitude of the response, τ is its average latency, and σ is its average width. The parameter σ is chosen so that $A \times \sigma$ is equal to the integral below the Gaussian function and corresponds to the average number of extra spikes per trial in response to the stimulus.

Classification analysis. To estimate how well a population of neurons ($n = 51$) can discriminate between two stimuli on a single trial, we used a template-matching algorithm¹⁵. For each pair of stimulus latencies (light stimulation at 32 and 62 ms latency, and at 32 and 92 ms latency), we aligned neuronal signals in two ways,

relative to the inhalation onset or relative to the stimulus onset. For every trial we built a response vector $\mathbf{r}_k = \{r_{1,1}, r_{1,2}, \dots, r_{1,m}, r_{2,1}, \dots, r_{2,m}, \dots, r_{n,m}\}$, where individual components, $r_{i,j}$, were number of spikes in a time bin j ($j = 1, \dots, m$) of a neuron i ($i = 1, \dots, n$). Time bins of 10 ms covered the interval from either the onset of inhalation or stimulus for the duration of 150 ms. For every stimulus and every trial, we estimated template vectors $\mathbf{r}_{k,s} = \langle \mathbf{r}_i \rangle_{i,i \neq k, S(i)=s}$ (where $S(i)$ is a stimulus type for a trial i , and $\langle \rangle_i$ is averaging over i), which averaged all trials for each stimuli $s = 1, 2$, excluding the given trial k . Then we assigned a given trial to one of the templates based on the shortest Euclidian distance between the response vector \mathbf{r}_k and the templates' vectors $\mathbf{r}_{k,s}$. The classification success was equal to a portion of correct assignments.

30. Bozza, T., McGann, J. P., Mombaerts, P. & Wachowiak, M. In vivo imaging of neuronal activity by targeted expression of a genetically encoded probe in the mouse. *Neuron* **42**, 9–21 (2004).
31. Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germ line. *Genes Dev.* **13**, 1524–1528 (1999).
32. Rinberg, D., Koulakov, A. & Gelperin, A. Sparse odor coding in awake behaving mice. *J. Neurosci.* **26**, 8857–8865 (2006).
33. Garthwaite, P. H., Jolliffe, I. T. & Jones, B. *Statistical Inference* (Oxford Univ. Press, 2002).

A high-resolution map of human evolutionary constraint using 29 mammals

Kerstin Lindblad-Toh^{1,2}, Manuel Garber^{1*}, Or Zuk^{1*}, Michael F. Lin^{1,3*}, Brian J. Parker^{4*}, Stefan Washietl^{3*}, Pouya Kheradpour^{1,3*}, Jason Ernst^{1,3*}, Gregory Jordan^{5*}, Evan Mauceli^{1*}, Lucas D. Ward^{1,3*}, Craig B. Lowe^{6,7,8*}, Alisha K. Holloway^{9*}, Michele Clamp^{1,10*}, Sante Gnerre^{1*}, Jessica Alföldi¹, Kathryn Beal⁵, Jean Chang¹, Hiram Clawson⁶, James Cuff¹¹, Federica Di Palma¹, Stephen Fitzgerald⁵, Paul Flicek⁵, Mitchell Guttman¹, Melissa J. Hubisz¹², David B. Jaffe¹, Irwin Jungreis³, W. James Kent⁹, Dennis Kostka⁹, Marcia Lara¹, Andre L. Martins¹², Tim Massingham⁵, Ida Moltke⁴, Brian J. Raney⁶, Matthew D. Rasmussen³, Jim Robinson¹, Alexander Stark¹³, Albert J. Vilella⁵, Jiayu Wen⁴, Xiaohui Xie¹, Michael C. Zody¹, Broad Institute Sequencing Platform and Whole Genome Assembly Team†, Kim C. Worley¹⁴, Christie L. Kovar¹⁴, Donna M. Muzny¹⁴, Richard A. Gibbs¹⁴, Baylor College of Medicine Human Genome Sequencing Center Sequencing Team†, Wesley C. Warren¹⁵, Elaine R. Mardis¹⁵, George M. Weinstock^{14,15}, Richard K. Wilson¹⁵, Genome Institute at Washington University†, Ewan Birney⁵, Elliott H. Margulies¹⁶, Javier Herrero⁵, Eric D. Green¹⁷, David Haussler^{6,8}, Adam Siepel¹², Nick Goldman⁵, Katherine S. Pollard^{9,18}, Jakob S. Pedersen^{4,19}, Eric S. Lander¹ & Manolis Kellis^{1,3}

The comparison of related genomes has emerged as a powerful lens for genome interpretation. Here we report the sequencing and comparative analysis of 29 eutherian genomes. We confirm that at least 5.5% of the human genome has undergone purifying selection, and locate constrained elements covering ~4.2% of the genome. We use evolutionary signatures and comparisons with experimental data sets to suggest candidate functions for ~60% of constrained bases. These elements reveal a small number of new coding exons, candidate stop codon readthrough events and over 10,000 regions of overlapping synonymous constraint within protein-coding exons. We find 220 candidate RNA structural families, and nearly a million elements overlapping potential promoter, enhancer and insulator regions. We report specific amino acid residues that have undergone positive selection, 280,000 non-coding elements exapted from mobile elements and more than 1,000 primate- and human-accelerated elements. Overlap with disease-associated variants indicates that our findings will be relevant for studies of human biology, health and disease.

A key goal in understanding the human genome is to discover and interpret all functional elements encoded within its sequence. Although only ~1.5% of the human genome encodes protein sequence¹, comparative analysis with the mouse², rat³ and dog⁴ genomes showed that at least 5% is under purifying selection and thus probably functional, of which ~3.5% consists of non-coding elements with probable regulatory roles. Detecting and interpreting these elements is particularly relevant to medicine, as loci identified in genome-wide association studies (GWAS) frequently lie in non-coding sequence⁵.

Although initial comparative mammalian studies could estimate the overall proportion of the genome under evolutionary constraint, they had little power to detect most of the constrained elements—especially the smaller ones. Thus, they focused only on the top 5% of constrained sequence, corresponding to less than ~0.2% of the genome^{4,6}. In 2005, we began an effort to generate sequence from a large collection of mammalian genomes with the specific goal of identifying and interpreting functional elements in the human genome on

the basis of their evolutionary signatures^{7,8}. Here we report our results to systematically characterize mammalian constraint using 29 eutherian (placental) genomes. We identify 4.2% of the human genome as constrained and ascribe potential function to ~60% of these bases using diverse lines of evidence for protein-coding, RNA, regulatory and chromatin roles, and we present evidence of exaptation and accelerated evolution. All data sets described here are publicly available in a comprehensive data set at the Broad Institute and University of California, Santa Cruz (UCSC).

Sequencing, assembly and alignment

We generated genome sequence assemblies for 29 mammalian species selected to achieve maximum divergence across the four major mammalian clades (Fig. 1a and Supplementary Text 1 and Supplementary Table 1). For nine species, we used genome assemblies based on ~7-fold coverage shotgun sequence, and for 20 species we generated ~2-fold coverage (2×), to maximize the number of species sequenced

¹Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, SE-751 23 Uppsala, Sweden. ³MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar St. Cambridge, Massachusetts 02139, USA. ⁴The Bioinformatics Centre, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. ⁵EMBL-EBI, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK. ⁶Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA. ⁷Department of Developmental Biology, Stanford University, Stanford, California 94305, USA. ⁸Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815, USA. ⁹Gladstone Institutes, University of California, 1650 Owens Street, San Francisco, California 94158, USA. ¹⁰BioTeam Inc, 7 Derosier Drive, Middleton, Massachusetts 01949, USA. ¹¹Research Computing, Division of Science, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ¹²Department of Biological Statistics & Computational Biology, Cornell University, Ithaca, New York 14853, USA. ¹³Research Institute of Molecular Pathology (IMP), A-1030 Vienna, Austria. ¹⁴Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ¹⁵Genome Institute at Washington University, Washington University School of Medicine, 4444 Forest Park Blvd., Saint Louis, Missouri 63108, USA. ¹⁶Genome Informatics Section, Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892 USA. ¹⁷NISC Comparative Sequencing Program, Genome Technology Branch and NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892 USA. ¹⁸Institute for Human Genetics, and Division of Biostatistics, University of California, 1650 Owens Street, San Francisco, California 94158, USA. ¹⁹Department of Molecular Medicine (MOMA), Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark.

*These authors contributed equally to this work.

†A full list of authors and their affiliations appears at the end of paper.

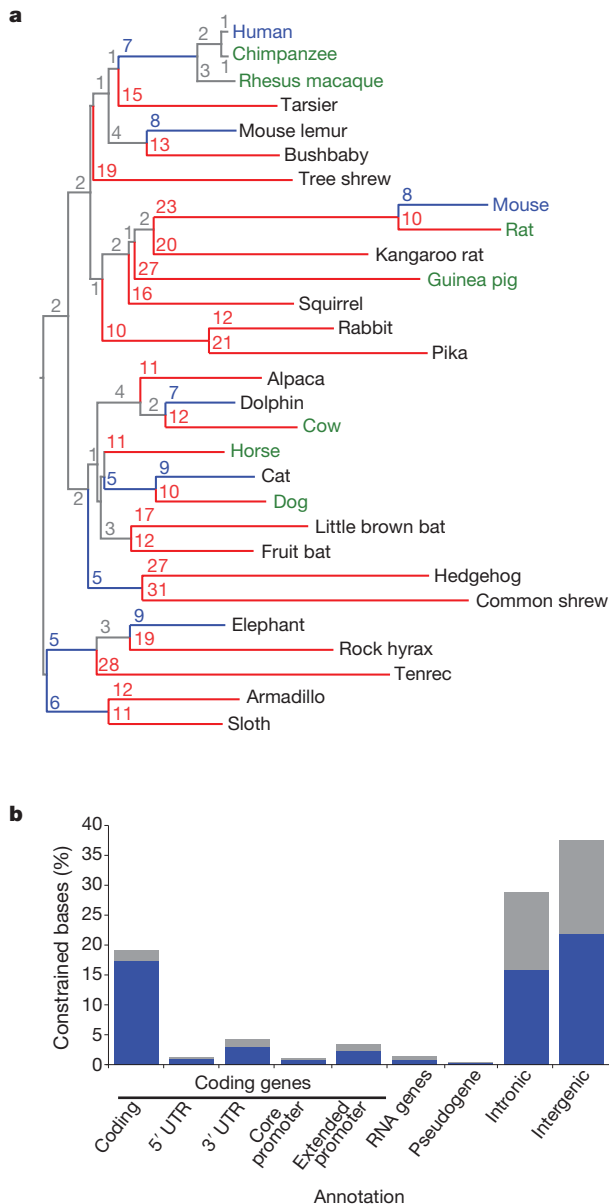


Figure 1 | Phylogeny and constrained elements from the 29 eutherian mammalian genome sequences. **a**, A phylogenetic tree of all 29 mammals used in this analysis based on the substitution rates in the MultiZ alignments. Organisms with finished genome sequences are indicated in blue, high quality drafts in green and 2x assemblies in black. Substitutions per 100 bp are given for each branch; branches with ≥ 10 substitutions are coloured red, blue indicates < 10 substitutions. **b**, At 10% FDR, 3.6 million constrained elements can be detected encompassing 4.2% of the genome, including a substantial fraction of newly detected bases (blue) compared to the union of the HMRD 50-bp + Siepel vertebrate elements¹⁷ (see Supplementary Fig. 4b for comparison to HMRD elements only). The largest fraction of constraint can be seen in coding exons, introns and intergenic regions. For unique counts, the analysis was performed hierarchically: coding exons, 5' UTRs, 3' UTRs, promoters, pseudogenes, non-coding RNAs, introns, intergenic. The constrained bases are particularly enriched in coding transcripts and their promoters (Supplementary Fig. 4c).

with available resources on capillary machines. Twenty genomes are first reported here, and nine were previously described (see Supplementary Information).

The power to detect constrained elements depends largely on the total branch length of the phylogenetic tree connecting the species⁹. The 29 mammals correspond to a total effective branch length of ~ 4.5

substitutions per site, compared to ~ 0.68 for the human-mouse-rat-dog comparison (HMRD), and thus should offer greater power to detect evolutionary constraint: the probability that a genomic sequence not under purifying selection will remain fixed across all 29 species is $P_1 < 0.02$ for single bases and $P_{12} < 10^{-25}$ for 12-nucleotide sequences, compared to $P_1 \sim 0.50$ and $P_{12} \sim 10^{-3}$ for HMRD.

For mammals for which we generated $2\times$ coverage, our assisted assembly approach¹⁰ resulted in a typical contig size $N50_c$ of 2.8 kb and a typical scaffold size $N50_s$ of 51.8 kb (Supplementary Text 2 and Supplementary Table 1) and high sequence accuracy (96% of bases had quality score Q20, corresponding to a $< 1\%$ error rate)¹¹. Compared to high-quality sequence across the 30 Mb of the ENCODE pilot project¹², we estimated average error rates of 1–3 miscalled bases per kilobase¹¹, which is ~ 50 -fold lower than the typical nucleotide sequence difference between the species, enabling high-confidence detection of evolutionary constraint (Supplementary Text 3).

We based our analysis on whole-genome alignments by MultiZ (Supplementary Text 4). The average number of aligned species was 20.9 at protein-coding positions in the human genome and 23.9 at the top 5% HMRD-conserved non-coding positions, with an average branch length of 4.3 substitutions per base in these regions (Supplementary Figs 1 and 2). In contrast, whole-genome average alignment depth is only 17.1 species with 2.9 substitutions per site, probably due to large deletions in non-functional regions⁴. The depth at ancestral repeats is 11.4 (Supplementary Fig. 1a), consistent with repeats being largely non-functional^{2,4}.

Detection of constrained sequence

Our analysis did not substantially change the estimate of the proportion of genome under selection. By comparing genome-wide conservation to that of ancestral repeats, we estimated the overall fraction of the genome under evolutionary constraint to be 5.36% at 50-bp windows (5.44% at 12-bp windows), using the SiPhy- ω statistic¹³, a measure of overall substitution rate (Supplementary Fig. 3), consistent with previous similar estimates^{2,4,14}. However, alternative methods^{15,16} and different ways of correcting for the varying alignment depths give higher estimates (see Supplementary Text 5 for details).

The additional species had a marked effect on our ability to identify the specific elements under constraint. With 29 mammals, we pinpoint 3.6 million elements spanning 4.2% of the genome, at a finer resolution of 12 bp (Fig. 1b and Supplementary Text 6, Supplementary Fig. 4, Supplementary Tables 2 and 3), compared to $< 0.1\%$ of the genome for HMRD 12-bp elements and 2.0% for HMRD 50-bp elements⁴. Elements previously detected using five vertebrates¹⁷ also detect a larger fraction of the genome ($\sim 4.1\%$), but only cover 45% of the mammalian elements detected here, suggesting that a large fraction of our elements are mammalian specific. The mean element size (36 bp) is considerably shorter than both previously detected HMRD elements (123 bp) and five-vertebrate elements (104 bp)¹⁷. For example, it is now possible to detect individual binding sites for the neuron-restrictive silencer factor (NRSF) in the promoter of the *NPAS4* gene, which are beyond detection power in previous data sets (Fig. 2 and Supplementary Fig. 5). We found a similar regional distribution of 12-bp elements (including the 2.6 million newly detected constrained elements) to previously detected HMRD elements ($r = 0.94$, Supplementary Fig. 6). Similar results were obtained with the PhastCons¹⁷ statistic (see Supplementary Text 6).

Using a new method, SiPhy- π , sensitive not just to the substitution rate but also to biases in the substitution pattern (for example, positions free to mutate between G and T only, Supplementary Fig. 7), we detected an additional 1.3% of the human genome in constrained elements (see Supplementary Tables 2 and 3). Most of the newly detected constrained nucleotides extend elements found by rate-based methods, but 22% of nucleotides lie in new elements (average length 17 bp) and are enriched in non-coding regions.

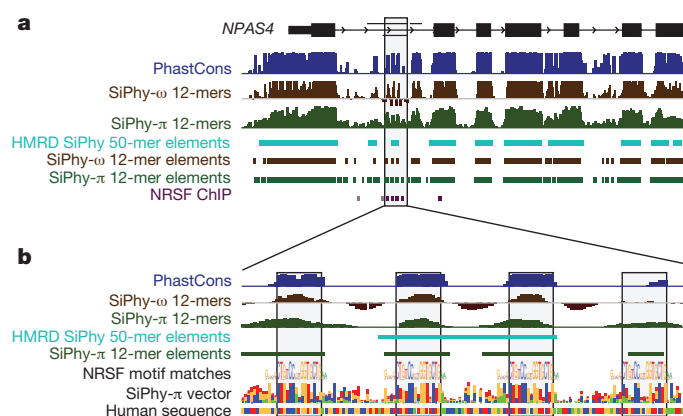


Figure 2 | Identification of four NRSF-binding sites in *NPAS4*. **a**, The neurological gene *NPAS4* has many constrained elements overlapping introns and the upstream intergenic region. The grey shaded box contained only one constrained element using HMRD, whereas analysis of 29 mammalian sequences reveals four smaller elements. **b**, These four constrained elements in the first intron correspond to binding sites for the NRSF transcription factor, known to regulate neuronal lineages.

Constraint within the human population

We observed that the evolutionary constraint acting on the 29 mammals is correlated with constraint within the human population, as assessed from human polymorphism data (Supplementary Text 7) and consistent with previous studies¹⁸. Mammalian constrained elements show a depletion in single-nucleotide polymorphisms (SNPs)¹⁹, and more constrained elements show even greater depletion. For example, in the top 1% most strongly conserved non-coding regions, SNPs occur at a 1.9-fold lower rate than the genome average, and the derived alleles have a lower frequency, consistent with purifying selection at many of these sites in the human population.

Moreover, at positions with biased substitution patterns across mammals, the observed human SNPs show a similar bias to the one observed across mammals (Supplementary Fig. 7). Thus, not only are constrained regions less likely to exhibit polymorphism in humans, but when such polymorphisms are observed, the derived alleles in humans tend to match the alleles present in non-human mammals, indicating a preference for the same alleles across both mammalian and human evolution.

Functional annotation of constraint

We first studied the overlap of the 3.6 million evolutionarily constrained elements ($\omega < 0.8$ and $P < 10^{-15}$) with known gene annotations (Fig. 1b). Roughly 30% of constrained elements were associated with protein-coding transcripts: ~25.3% overlap mature messenger RNAs (including 19.6% in coding exons, 1.2% in 5' untranslated regions (5' UTRs) and 4.4% in 3' UTRs), and an additional 4.4% reside within 2 kb of transcriptional start sites (1.2% reside within 200 bases).

The majority of constrained elements, however, reside in intronic and intergenic regions (29.7% and 38.6%, respectively). To study their biological roles and provide potential starting points to understand these large and mostly uncharted territories, we next studied their overlap with evolutionary signatures^{7,8,20,21} characteristic of specific types of features and a growing collection of public large-scale experimental data.

Protein-coding genes and exons

Despite intense efforts to annotate protein-coding genes over the past decade^{20,22–24}, we detected 3,788 candidate new exons (a 2% increase) using evolutionary signatures characteristic of protein-coding exons²⁵. Of these, 54% reside outside transcripts of protein-coding genes, 19% within introns, and 13% in UTRs of known coding genes (Supplementary Text 8, Supplementary Tables 4 and 5). Our methods recovered

92% of known coding exons that were larger than 10 codons and fall in syntenic regions, the remainder showing non-consensus splice sites, unusual features, or poor conservation.

The majority of new exon candidates ($>58\%$) are supported by evidence of transcription measured in 16 human tissues²⁶ (Supplementary Fig. 8a) or similarity to known Pfam protein domains. Thirty-one per cent of intronic and 13% of intergenic predictions extend known transcripts, and 5% and 11% respectively reside in new transcript models. The newly detected exons are more tissue specific than known exons (mean of 3 tissues versus 12) and are expressed at fivefold lower levels. Directed experiments and manual curation will be required to complete the annotation of the few hundred protein-coding genes that probably remain unannotated²⁷.

We found apparent stop codon readthrough²⁸ of four genes based on continued protein-coding constraint after an initial conserved stop codon²⁹ and until a subsequent stop codon (Supplementary Text 9 and Supplementary Fig. 8b). Readthrough in *SACMIL* could be triggered by an 80-base conserved RNA stem loop predicted by RNAz³⁰, lying four bases downstream of the readthrough stop codon.

We also detected coding regions with a very low synonymous substitution rate, indicating additional sequence constraints beyond the amino acid level (Supplementary Text 9). We found $>10,000$ such synonymous constraint elements (SCEs) in more than one-quarter of all human genes³¹. Initial analysis indicates potential roles in splicing regulation (34% span an exon–exon junction), A-to-I editing, microRNA (miRNA) targeting and developmental regulation. *HOX* genes contain several top candidates (Fig. 3a), including two previously validated developmental enhancers^{32,33}.

RNA structures and families of structural elements

We next used evolutionary signatures characteristic of conserved RNA secondary structures³⁴ to reveal 37,381 candidate structural elements (Supplementary Text 10 and Supplementary Fig. 9a), covering ~1% of constrained regions. For example, the *XIST* large intergenic non-coding RNA (lincRNA), known to bind chromatin and enable X-chromosome inactivation³⁵, contains a newly predicted structure in its 3' end (Supplementary Fig. 9b, f)—distinct from other known structures³⁶—which seems to be the source of chromatin-associated short RNAs³⁷.

Sequence- and structure-based clustering of predictions outside protein-coding exons revealed 1,192 novel families of structural RNAs (Supplementary Text 10). We focused on a high-scoring subset consisting of 220 families with 725 instances, which also showed the highest thermodynamic stability³⁰ (Supplementary Figs 9a and 10), DNase hypersensitivity, expression pattern correlation across tissues and intergenic expression enrichment (Supplementary Fig. 9a). We also expanded both known and novel families by including additional members detected by homology to existing members.

Noteworthy examples include: a glycyl-tRNA family, including a new member in *POPI*, involved in tRNA maturation and probably involved in feedback regulation of *POPI*; three intronic families of long hairpins in ion-channel genes known to undergo A-to-I RNA editing and possibly involved in regulation of the editing event; an additional member of a family of 5' UTR hairpins overlapping the start codon of collagen genes and potential new miRNA genes that extend existing families³⁷.

Two of the largest novel families consist of short AU-rich hairpins of 6–7 bp that share the same strong consensus motif in their stem. These occur in the 3' UTRs of genes in several inflammatory response pathways, the post-transcriptional regulation of which often involves AU-rich elements (AREs). Indeed, two homologous hairpins in *TNF* and *CSF3* correspond to known mRNA-destabilization elements, suggesting roles in mRNA stability for the two families³⁷.

Lastly, a family of six conserved hairpin structures (Supplementary Fig. 9g) was found in the 3' UTR of the *MAT2A* gene³⁷, which is involved in the synthesis of S-adenosylmethionine (SAM), the primary

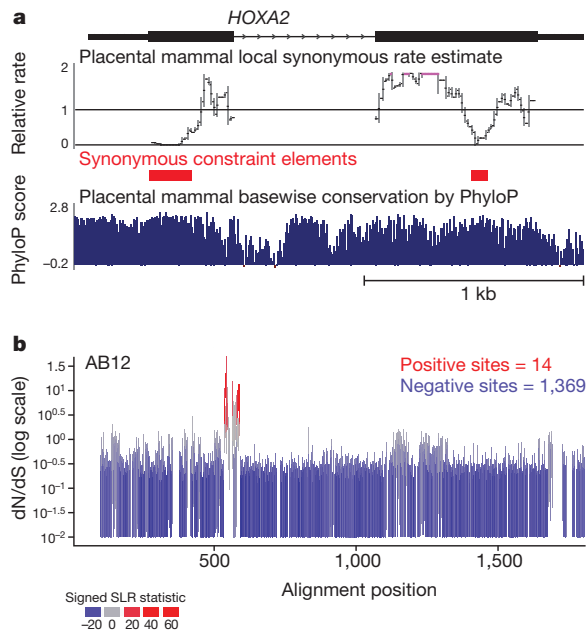


Figure 3 | Examination of evolutionary signatures identifies SCEs and evidence of positive selection. **a**, Two regions within the *HOXA2* open reading frame are identified as SCEs (red), corresponding to overlapping functional elements within coding regions. Note that the synonymous rate reductions are not obvious from the base-wise conservation measure (in blue). Both elements have been characterized as enhancers driving *HOXA2* expression in distinct segments of the developing mouse hindbrain. The element in the first exon encodes Hox–Pbx-binding sites and drives expression in rhombomere 4 (ref. 33), whereas the element in the second exon contains Sox-binding sites and drives expression in rhombomere 2 (ref. 32). Synonymous constraint elements are also found in most other *HOX* genes, and up to a quarter of all genes. **b**, Although ~85% of genes show only negative (purifying) selection and 9% of genes show uniform positive selection, the remaining 6% of genes, including *AB12*, show only localized regions of positively selected sites. Each vertical bar covers the estimated 95% confidence interval for dN/dS at that site (with values of 0 truncated to 0.01 to accommodate the log scaling), and bars are coloured according to a signed version of the SLR statistic for non-neutral evolution: blue for sites under purifying selection, grey for neutral sites and red for sites under positive selection.

methyl donor in human cells. All six hairpins consist of a 12–18-bp stem and a 14-bp loop region with a deeply conserved sequence motif (Supplementary Fig. 9e), and may be involved in sensing SAM concentrations, which are known to affect *MAT2A* mRNA stability³⁸.

Conservation patterns in promoters

As different types of conservation in promoters may imply distinct biological functions³⁹, we classified the patterns of conservation within core promoters into three categories: (1) those with uniformly ‘high’ constraint (7,635 genes, 13,996 transcripts); (2) uniformly ‘low’ constraint (2,879 genes, 4,135 transcripts); and (3) ‘intermittent’ constraint, consisting of alternating peaks and troughs of conservation (14,271 genes and 29,814 transcripts) (Supplementary Fig. 11a). High and intermittent constraint promoters are both associated with CpG islands (~66%), whereas low constraint promoters have significantly lower overlap (~41%), and all three classes show similar overlap with functional TATA boxes (2–3%, see Supplementary Text 11).

These groups show distinct Gene Ontology enrichments (Supplementary Fig. 11b), with high-constraint promoters involved in development (P with Bonferroni correction ($P_{\text{Bonf}} < 10^{-30}$), intermittent constraint in basic cellular functions ($P_{\text{Bonf}} < 5 \times 10^{-4}$), and low-constraint promoters in immunity, reproduction and perception, functions expected to be under positive selection and lineage-specific adaptation².

High constraint may reflect cooperative binding of many densely binding factors, as previously suggested for developmental genes⁶. Intermittent constraint promoters, the peak-spacing distribution of which was suggestive of the periodicity of the DNA helix turns, may reflect loosely interacting factors (Supplementary Fig. 11c, d). Low constraint may reflect rapid motif turnover, under neutral drift or positive selection.

Identifying specific instances of regulatory motifs

Data from just four species (HMRD) was sufficient to create a catalogue of known and novel motifs with many conserved instances across the genome²¹. The power to discover such motifs was high, because one can aggregate data across hundreds of motif instances. Not surprisingly, the additional genomes therefore had little effect on the ability to discover new motifs (known motifs showed 99% correlation in genome-wide motif conservation scores, Supplementary Figs 12 and 13).

In contrast, the 29 mammalian genomes markedly improved our ability to detect individual motif instances, making it possible to predict specific target sites for 688 regulatory motifs corresponding to 345 transcription factors (Supplementary Fig. 14). We chose to identify motif instances at a false discovery rate (FDR) of 60%, representing a reasonable compromise between specificity and sensitivity given the available discovery power (Supplementary Text 12), and matching the experimental specificity of chromatin immunoprecipitation (ChIP) experiments for identifying biologically significant targets⁴⁰. Higher levels of stringency could be obtained by sequencing additional species.

We identified 2.7 million conserved instances (Supplementary Table 6), enabling the construction of a regulatory network linking 375 motifs to predicted targets, with a median of 21 predicted regulators per target gene (25th percentile, 10; 75th percentile, 39). The number of target sites (average, 4,277; 25th percentile, 1,407; 75th percentile, 10,782) are comparable to those found in ChIP experiments, and have the advantage that they are detected at nucleotide resolution, enabling us to use them to interpret disease-associated variants for potential regulatory functions. However, some motifs never reached high confidence values, and others did so at very few instances.

The motif-based targets show strong agreement with experimentally defined binding sites from ChIP experiments (Supplementary Table 7). For long and distinct motifs, such as CTCF and NRSF, the fraction of instances overlapping experimentally observed binding matches the fraction predicted by the confidence score (for example, at 80% confidence 70% of NRSF motif instances overlapped bound sites, and at ~50% confidence 40% overlapped), despite potential confounding aspects such as condition-specific binding, overlapping motifs between factors, or non-specific binding. Moreover, increasing confidence levels showed increasing overlap with experimental binding (Supplementary Figs 14–16). For example, YY1 enrichment for bound sites increased from 42-fold to 168-fold by focusing on conserved instances. Lastly, combining motif conservation and experimental binding led to increased enrichment for candidate tissue-specific enhancers, suggesting that the two provide complementary information. Within bound regions, the evolutionary signal reveals specific motif instances with high precision (for example, Figs 2 and 4 and Supplementary Fig. 17).

Chromatin signatures

To suggest potential functions for the ~68% of ‘unexplained’ constrained elements outside coding regions, UTRs or proximal promoters, we used chromatin state maps from CD4 T cells⁴¹ (Supplementary Fig. 18) and nine diverse cell types⁴² (Supplementary Text 13 and Supplementary Fig. 19). In T cells, constrained elements were most enriched for promoter-associated states (up to fivefold), an insulator state and a specific repressed state (2.2-fold), and numerous enhancer

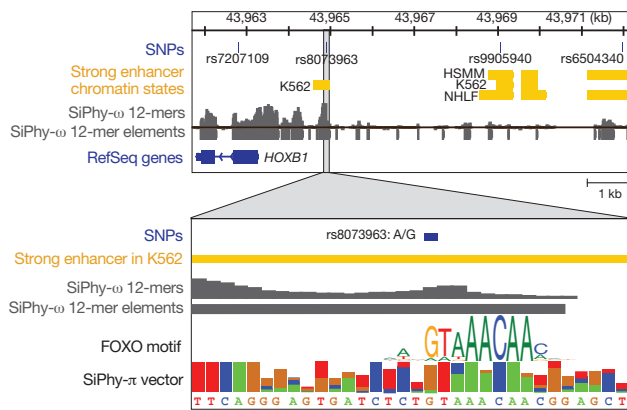


Figure 4 | Using constraint to identify candidate mutations. Conservation can help us resolve amid multiple SNPs the ones that disrupt conserved functional elements and are likely to have regulatory roles. In this example, a SNP (rs6504340) associated with tooth development is strongly linked to a conserved intergenic SNP, rs8073963, 7.1 kb away, which disrupts a deeply conserved Forkhead-family motif in a strong enhancer. Although the SNPs shown here stem from GWAS on HapMap data, the same principle should be applicable to associated variants detected by resequencing the region of interest.

states (1.5–2-fold), together covering 7.1% of the unexplained elements at 2.1-fold enrichment. In the nine cell types, enriched promoter, enhancer and insulator states cover 36% of unexplained elements at ~1.75-fold enrichment, with locations active in multiple cell types showing even stronger enrichment (Supplementary Fig. 20).

Overall, chromatin states indicate possible functions (at 1.74-fold enrichment) for 37.5% ($N = 987,985$) of unexplained conserved elements (27% of all conserved elements), suggesting meaningful association for at least 16% of unexplained constrained bases. Although current experiments only provide nucleosome-scale (~200-bp) resolution, we expect higher-resolution experimental assays that more precisely pinpoint regulatory regions to show further increases in enrichment. The increased overlap observed with additional cell types suggests that new cell types will help elucidate additional elements. Of course, further experimental tests will be required to validate the predicted functional roles.

Accounting for constrained elements

Overall, ~30% of constrained elements overlap were associated with protein-coding transcripts, ~27% overlap specific enriched chromatin states, ~1.5% novel RNA structures, and ~3% conserved regulatory motif instances (Supplementary Text 13, 14). Together, ~60% of constrained elements overlap one of these features, with enrichments ranging from 1.75-fold for chromatin states (compared to unannotated regions) up to 17-fold for protein-coding exons (compared to the whole genome).

Implications for interpreting disease-associated variants

In the non-protein-coding genome, SNPs associated with human diseases in genome-wide association studies are 1.37-fold enriched for constrained regions, relative to HapMap SNPs (Supplementary Text 15 and Supplementary Table 8). This is notable because only a small proportion of the associated SNPs are likely to be causative, whereas the rest are merely in linkage disequilibrium with causative variants.

Accordingly, constrained elements should be valuable in focusing the search for causative variants among multiple variants in linkage disequilibrium. For example, in an intergenic region between *HOXB1* and *HOXB2* associated with tooth development phenotypes⁴³, the reported SNP (rs6504340) is not conserved, but a linked SNP (rs8073963) sits in a constrained element 7.1 kb away. Moreover, rs8073963 disrupts a deeply conserved FOXO2 motif instance within a predicted enhancer (Fig. 4), making it a candidate mutation for

further follow-up. Similar examples of candidate causal variants are found for diverse phenotypes such as height or multiple sclerosis, and similar analyses could be applied to case-control resequencing data.

Evolution of constrained elements

We next sought to identify signatures of positive selection that may accompany functional adaptations of different species to diverse environments and new ecosystems.

Codon-specific selection

We used the ratio dN/dS of non-synonymous to synonymous codon substitutions as evidence of positive selection (>1) or negative selection (<1). Although dN/dS is typically calculated for whole genes, the additional mammals sequenced enabled analysis at the codon level: simulations predicted a 250-fold gain in sensitivity compared to HMRD, identifying 53% of positive sites at 5% FDR (Supplementary Text 16).

Applying this test to 6.05 million codons in 12,871 gene trees, we found evidence of strong purifying selection ($dN/dS < 0.5$) for 84.2% of codons and positive selection ($dN/dS > 1.5$) for 2.4% of codons (with 94.1% of sites <1 and 5.9% >1 ; Supplementary Table 9). At 5% FDR, we found 15,383 positively selected sites in 4,431 proteins. The genes fall into three classes based on the distribution of selective constraint: 84.8% of genes show uniformly high purifying selection, 8.9% show distributed positive selection across their length and 6.3% show localized positive selection concentrated in small clusters (Fig. 3b and Supplementary Fig. 21, Supplementary Tables 10 and 11).

Genes with distributed positive selection were enriched in such functional categories as immune response ($P_{\text{Bonf}} < 10^{-16}$) and taste perception ($P_{\text{Bonf}} < 10^{-10}$), which are known to evolve rapidly, but also in some unexpected functions such as meiotic chromosome segregation ($P_{\text{Bonf}} < 10^{-23}$) and DNA-dependent regulation of transcription ($P_{\text{Bonf}} < 10^{-19}$; Supplementary Table 12). Localized positive selection was enriched in core biochemical processes, including microtubule-based movement ($P_{\text{Bonf}} < 10^{-10}$), DNA topological change ($P_{\text{Bonf}} < 10^{-4}$) and telomere maintenance ($P_{\text{Bonf}} < 7 \times 10^{-3}$), suggesting adaptation at important functional sites.

Focusing on 451 unique Pfam protein-domain annotations, we found abundant purifying selection, with 225 domains showing purifying selection for $>75\%$ of their sites, and 447 domains showing negative selection for $>50\%$ of their sites (Supplementary Table 13). Domains with substantial fractions of positively selected sites include CRAL/TRIO involved in retinal binding (2.6%), proteinase-inhibitor-cystatin involved in bone remodelling (2.2%) and the secretion-related EMP24/GOLD/P24 family (1.6%).

Exaptation of mobile elements

Mobile elements provide an elegant mechanism for distributing a common sequence across the genome, which can then be retained in locations where it confers advantageous regulatory functions to the host—a process termed exaptation. Our data revealed $>280,000$ mobile element exaptations common to mammalian genomes covering ~7 Mb (Supplementary Text 17), a considerable expansion from the ~10,000 previously recognized cases⁴⁴. Of the ~1.1 million constrained elements that arose during the 90 million years between the divergence from marsupials and the eutherian radiation, we can trace $>19\%$ to mobile element exaptations. Often only a small fraction (median ~11%) of each mobile element is constrained, in some cases matching known regulatory motifs. Recent exaptations are generally found near ancestral regulatory elements, except in gene deserts, which are abundant in ancestral elements but show few recent exaptations ($P < 10^{-300}$, Supplementary Fig. 22).

Accelerated evolution in the primate lineage

Lineage-specific rapid evolution in ancestrally constrained elements previously revealed human positive selection associated with brain

and limb development⁴⁵. Applying this signature to the human and primate lineages, we identified 563 human-accelerated regions (HARs) and 577 primate-accelerated regions (PARs) at FDR <10% (Supplementary Text 18, Supplementary Tables 14 and 15), significantly expanding the 202 previously known HARs⁴⁶. Fifty-four HARs (9.4%) and 49 PARs (8.5%) overlap enhancer-associated chromatin marks and experimentally validated enhancers (Supplementary Text 18). Substitution patterns in HARs suggest that GC-biased gene conversion (BGC) is not responsible for the accelerated evolution in the vast majority of these regions (~15% show evidence of BGC).

Genes harbouring or neighbouring HARs and PARs are enriched for extracellular signalling, receptor activity, immunity, axon guidance, cartilage development and embryonic pattern specification (Supplementary Fig. 23). For example, the *FGF13* locus associated with an X-linked form of mental retardation contains four HARs near the 5' ends of alternatively spliced isoforms of *FGF13* expressed in the nervous system, epithelial tissues and tumours, suggesting human-specific changes in isoform regulation (Supplementary Fig. 24).

Discussion

Comparative analysis of 29 mammalian genomes reveals a high-resolution map of >3.5 million constrained elements that encompass ~4% of the human genome and suggest potential functional classes for ~60% of the constrained bases; the remaining 40% show no overlap and remain uncharacterized. We report previously undetected exons and overlapping functional elements within protein-coding sequence, new classes of RNA structures, promoter conservation profiles and predicted targets of transcriptional regulators. We also provide evidence of evolutionary innovation, including codon-specific positive selection, mobile element exaptation and accelerated evolution in the primate and human lineages.

By focusing our comparison on only eutherian mammals, we discover functional elements relevant to this clade, including recent eutherian innovations. This is especially important for discovering regulatory elements, which can be subject to rapid turnover⁴⁷. Indeed, a previous comparison indicated that only 80% of 50-bp non-coding elements are shared with opossum⁴⁸, and the current 12-bp analysis shows ~64% of non-coding elements shared with opossum, and only 6% with stickleback fish. Many eutherian elements are thus probably missing from previous maps of vertebrate constraint¹⁷.

Sequencing of additional species should enable discovery of lineage-specific elements within mammalian clades, and provide increased resolution for shared mammalian constraint. We estimate that 100–200 eutherian mammals (15–25 neutral substitutions per site) will enable single-nucleotide resolution. The majority of this branch length is present within the Laurasiatherian and Euarchontoglires branches, which also contain multiple model organisms. These are ideal next targets for sequencing as part of the Genome 10K effort⁴⁹, aiming to sequence 10,000 vertebrate species. Within the primate clade, a branch length of ~1.5 could probably be achieved, enabling primate-specific selection studies, albeit at lower resolution. Lastly, human-specific selection should be detectable by combining data across genomic regions and by comparing thousands of humans⁵⁰.

The constrained elements reported here can be used to prioritize disease-associated variants for subsequent study, providing a powerful lens for elucidating functional elements in the human genome complementary to ongoing large-scale experimental endeavours such as ENCODE and Roadmap Epigenomics. Experimental studies require prior knowledge of the biochemical activity sought and reveal regions active in specific cell types and conditions. Comparative approaches provide an unbiased catalogue of shared functional regions independent of biochemical activity or condition, and thus can capture experimentally intractable or rare activity patterns. With increasing branch length, they can provide information on ancestral and recent selective pressures across mammalian clades and within

the human population. Ultimately, the combination of disease genetics, comparative and population genomics and biochemical studies have important implications for understanding human biology, health and disease.

METHODS SUMMARY

A full description of materials and methods, including sample selection and sequencing strategy, assembly strategies and results, error estimation and correction, alignment details, estimation of genome portion under constraint, detection of constrained elements, mammalian constraint versus human polymorphism, protein coding genes, detection of stop codon readthrough and synonymous constraint elements, RNA structure detection, patterns of promoter constraint, regulatory motif discovery, correlation with chromatin state information, overall accounting of constraint elements, comparison with disease-associated variants, detection of codon-specific positive selection, exaptation of ancestral repeat elements, and human and primate accelerated regions is available in Supplementary Information. All animal experiments were approved by the MIT Committee for Animal Care.

Received 25 January; accepted 5 September 2011.

Published online 12 October 2011.

- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
- Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
- Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Stark, A. *et al.* Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219–232 (2007).
- Cooper, G. M., Brudno, M., Green, E. D., Batzoglu, S. & Sidow, A. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res.* **13**, 813–820 (2003).
- Gnerre, S., Lander, E. S., Lindblad-Toh, K. & Jaffe, D. B. Assisted assembly: how to improve a *de novo* genome assembly by using related species. *Genome Biol.* **10**, R88 (2009).
- Hubisz, M. J., Lin, M. F., Kellis, M. & Siepel, A. Error and error mitigation in low-coverage genome assemblies. *PLoS ONE* **14**, e17034 (2011).
- Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
- Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54–i62 (2009).
- Chiaromonte, F. *et al.* The share of human genomic DNA under selection estimated from human-mouse genomic alignments. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 245–254 (2003).
- Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comput. Biol.* **6**, e1001025 (2010).
- Meader, S., Ponting, C. P. & Lunter, G. Massive turnover of functional sequence in human and other mammalian genomes. *Genome Res.* **20**, 1335–1343 (2010).
- Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
- Drake, J. A. *et al.* Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nature Genet.* **38**, 223–227 (2006).
- Keinan, A., Mullikin, J. C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genet.* **39**, 1251–1255 (2007).
- Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl Acad. Sci. USA* **104**, 19428–19433 (2007).
- Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Siepel, A. *et al.* Targeted discovery of novel human exons by comparative genomics. *Genome Res.* **17**, 1763–1773 (2007).
- Pruitt, K. D. *et al.* The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* **19**, 1316–1323 (2009).
- Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein-coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2010).
- Guttman, M. *et al.* *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnol.* **28**, 503–510 (2010).

27. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (suppl. 1), 1–9 (2006).
28. Lin, M. F. *et al.* Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* **17**, 1823–1836 (2007).
29. Jungreis, I. *et al.* Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. *Genome Res.* doi:10.1101/gr.119974.110 (in the press).
30. Washietl, S., Hofacker, I. L. & Stadler, P. F. Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA* **102**, 2454–2459 (2005).
31. Lin, M. F. *et al.* Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res.* doi:10.1101/gr.108753.110 (in the press).
32. Tumpel, S., Cambronero, F., Sims, C., Krumlauf, R. & Wiedemann, L. M. A regulatory module embedded in the coding region of *Hoxa2* controls expression in rhombomere 2. *Proc. Natl Acad. Sci. USA* **105**, 20077–20082 (2008).
33. Lampe, X. *et al.* An ultraconserved Hox–Pbx responsive element resides in the coding sequence of *Hoxa2* and is active in rhombomere 4. *Nucleic Acids Res.* **36**, 3214–3225 (2008).
34. Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLOS Comput. Biol.* **2**, e33 (2006).
35. Lee, J. T. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* **23**, 1831–1842 (2009).
36. Maenner, S. *et al.* 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* **8**, e1000276 (2010).
37. Parker, B. J. *et al.* New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. *Genome Res.* doi:10.1101/gr.112516.110 (in the press).
38. Martinez-Chantar, M. L. *et al.* L-methionine availability regulates expression of the methionine adenosyltransferase 2A gene in human hepatocarcinoma cells: role of S-adenosylmethionine. *J. Biol. Chem.* **278**, 19885–19890 (2003).
39. Baek, D., Davis, C., Ewing, B., Gordon, D. & Green, P. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res.* **17**, 145–155 (2007).
40. Kheradpour, P., Stark, A., Roy, S. & Kellis, M. Reliable prediction of regulator targets using 12 *Drosophila* genomes. *Genome Res.* **17**, 1919–1931 (2007).
41. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnol.* **28**, 817–825 (2010).
42. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
43. Pillas, D. *et al.* Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. *PLoS Genet.* **6**, e1000856 (2010).
44. Lowe, C. B., Bejerano, G. & Haussler, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl Acad. Sci. USA* **104**, 8005–8010 (2007).
45. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
46. Pollard, K. S. *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168 (2006).
47. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
48. Mikkelsen, T. S. *et al.* Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
49. Genome 10K Community Of Scientists. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J. Hered.* **100**, 659–674 (2009).
50. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank O. Ryder, E. Fuchs, D. Haring, A. Walsh, D. Duffield, S. Wong, T. Alvarado, J. Boylan, S. Combes, P. deJong, J. Allman, J. Patton, D. McMullen, D. Hafner, D. Miller, T. Kunz, G. Hewitt, J. Searle, H. Künzle and D. Williams for providing organismal material. We thank L. Gaffney for help with figures. This work was supported by the National Human Genome Research Institute (NHGRI), including grant U54 HG003273 (R.A.G.), National Institute for General Medicine (NIGMS) grant no.

GM82901 (Pollard laboratory) and the European Science Foundation (EURYL award to K.L.-T.), NSF National Science Foundation (NSF) postdoctoral fellowship award 0905968 (J.E.), National Science Foundation CAREER 0644282 and NIH R01 HG004037 and the Sloan Foundation (M.K.), and an Erwin Schrödinger Fellowship of the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (S.W.), the Gates Cambridge Trust (G.J.), Novo Nordisk Foundation (B.J.P. and J.W.); a Statistics Network Fellowship, Department of Mathematical Sciences, University of Copenhagen (B.J.P.); the David and Lucile Packard Foundation (A.S.); the Danish Council for Independent Research Medical Sciences (J.S.P.); The Lundbeck Foundation (J.S.P.).

Author Contributions K.L.-T., E.S.L. and M.K. led the project and oversaw the analysis. K.L.-T., M.C., J.Ch., E.H.M., E.D.G. and E.S.L. planned the project. K.L.-T., F.D.P., M.L., E.S.L., K.C.W., C.L.K., D.M.M., R.A.G., W.C.W., E.R.M., G.M.W. and R.K.W. oversaw or significantly contributed to data generation. S.G. assembled the 2 × genomes. Major contributions to analysis were made by M.Ga., O.Z. and M.C. to evaluate measures and patterns of evolutionary selection, M.F.L. to evaluate protein-coding potential and translational readthrough, B.J.P., S.W. to analyse RNA structures and families, P.K. on regulatory motifs and motif instances, J.E. on chromatin states, G.J. on codon-specific positive selection, E.M. on promoter motif, L.D.W. on GWAS overlap with conserved elements, C.B.L. on exaptation and A.K.H., K.S.P. on HARs and PARs. J.A., K.B., H.C., J.Cu., S.F., P.F., M.Gu., M.J.H., D.B.J., I.J., W.J.K., D.K., A.L.M., T.M., I.M., B.J.R., M.D.R., J.R., A.St., A.J.V., J.W., X.X., M.C.Z., E.B., E.H.M., J.H., D.H., A.Si., N.G. and J.S.P. performed or oversaw various analyses. K.L.-T., E.S.L. and M.K. wrote the paper with input from the other authors.

Author Information A complete set of data files can be downloaded from the Broad website (<https://www.broadinstitute.org/scientific-community/science/projects/mammals-models/29-mammals-project-supplementary-info>) or viewed using the UCSC Genome Browser (<http://genome.ucsc.edu/index.php/29mammals>) or the Broad Institute Integrative Genome Viewer (<http://www.broadinstitute.org/igv/projects/29mammals>). NCBI accession numbers for all newly sequenced genomes can be found in Supplementary Table 1. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.L.-T. (kersli@broadinstitute.org), E.S.L. (lander@broadinstitute.org) or M.K. (manoli@mit.edu).

Broad Institute Sequencing Platform and Whole Genome Assembly Team

Jen Baldwin¹, Toby Bloom¹, Chee Whye Chin¹, Dave Heiman¹, Robert Nicol¹, Chad Nusbaum¹, Sarah Young¹ & Jane Wilkinson¹

Baylor College of Medicine Human Genome Sequencing Center Sequencing Team

Andrew Cree², Huyen H. Dinh², Gerald Fowler², Shalili Jhangiani², Vandita Joshi², Sandra Lee², Lora R. Lewis², Lynne V. Nazareth², Geoffrey Okwuonu² & Jireh Santibanez²

Genome Institute at Washington University

Kim Delehaunty³, David Dooling³, Catrina Fronik³, Lucinda Fulton³, Bob Fulton³, Tina Graves³, Patrick Minx³ & Erica Sodergren^{3,4}

¹Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), 320 Charles Street, Cambridge, Massachusetts 02142, USA. ²Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. ³Genome Institute at Washington University, Washington University School of Medicine, 4444 Forest Park Blvd, Saint Louis, Missouri 63108, USA. ⁴Research Institute of Molecular Pathology (IMP), A-1030 Vienna, Austria.

Genome sequencing reveals insights into physiology and longevity of the naked mole rat

Eun Bae Kim^{1*}, Xiaodong Fang^{2*}, Alexey A. Fushan^{1*}, Zhiyong Huang^{2*}, Alexei V. Lobanov³, Lijuan Han², Stefano M. Marino³, Xiaoqing Sun², Anton A. Turanov³, Pengcheng Yang², Sun Hee Yim³, Xiang Zhao², Marina V. Kasaikina³, Nina Stoletzki³, Chunfang Peng², Paz Polak³, Zhiqiang Xiong², Adam Kiezun³, Yabing Zhu², Yuanxin Chen², Gregory V. Kryukov^{3,4}, Qiang Zhang², Leonid Peshkin⁵, Lan Yang², Roderick T. Bronson⁶, Rochelle Buffenstein⁷, Bo Wang², Changlei Han², Qiye Li², Li Chen², Wei Zhao², Shamil R. Sunyaev^{3,4}, Thomas J. Park⁸, Guojie Zhang², Jun Wang^{2,9,10} & Vadim N. Gladyshev^{1,3,4}

The naked mole rat (*Heterocephalus glaber*) is a strictly subterranean, extraordinarily long-lived eusocial mammal¹. Although it is the size of a mouse, its maximum lifespan exceeds 30 years, making this animal the longest-living rodent. Naked mole rats show negligible senescence, no age-related increase in mortality, and high fecundity until death². In addition to delayed ageing, they are resistant to both spontaneous cancer and experimentally induced tumorigenesis^{3,4}. Naked mole rats pose a challenge to the theories that link ageing, cancer and redox homeostasis. Although characterized by significant oxidative stress⁵, the naked mole rat proteome does not show age-related susceptibility to oxidative damage or increased ubiquitination⁶. Naked mole rats naturally reside in large colonies with a single breeding female, the 'queen', who suppresses the sexual maturity of her subordinates⁷. They also live in full darkness, at low oxygen and high carbon dioxide concentrations⁸, and are unable to sustain thermogenesis⁹ nor feel certain types of pain^{10,11}. Here we report the sequencing and analysis of the naked mole rat genome, which reveals unique genome features and molecular adaptations consistent with cancer resistance, poikilothermy, hairlessness and insensitivity to low oxygen, and altered visual function, circadian rhythms and taste sensing. This information provides insights into the naked mole rat's exceptional longevity and ability to live in hostile conditions, in the dark and at low oxygen. The extreme traits of the naked mole rat, together with the reported genome and transcriptome information, offer opportunities for understanding ageing and advancing other areas of biological and biomedical research.

We applied a whole-genome shotgun strategy to sequence the genome of an individual male naked mole rat (NMR) (Table 1 and Supplementary Tables 1–3). The sequencing depth of 98.6% of the

genome assembly was more than 20-fold (Supplementary Figs 1–4). The mitochondrial genome was also assembled. Approximately 25% of the NMR genome was represented by transposon-derived repeats, which is lower than in other mammals (40% in human, 37% in mouse, and 35% in rat genomes) (Supplementary Tables 4 and 5, Supplementary Figs 5–7). The predicted NMR gene set included 22,561 genes (Table 1 and Supplementary Table 6), which is comparable to other mammals (22,389 in human, 23,317 in mouse, and 22,841 in rat). Of these, 21,394 (94.8%) genes were transcribed (based on the RNA-seq data for seven organs). More than 98% of NMR genes could be functionally annotated using homology approaches (Supplementary Table 7), and the quality of predicted genes was comparable to that of well-annotated mammalian genomes (Supplementary Tables 6 and 8 and Supplementary Fig. 8).

Most of the NMR genome (93%) showed synteny to human, mouse or rat genomes (Supplementary Table 9), and pairwise comparisons suggested a relatively low rate of NMR genome rearrangements after the split from the murid common ancestor. We defined common synteny blocks in human, mouse, rat and NMR genomes and identified segmental duplications and lineage-specific insertions and deletions (Supplementary Tables 10 and 11 and Supplementary Fig. 9). By analysing single-copy orthologous groups, we constructed a phylogenetic tree involving the NMR and other mammals (Fig. 1). As expected, the NMR placed within rodents and its ancestor split from the ancestor of rats and mice approximately 73 million years ago, whereas the ancestor of NMR, mouse and rat split from rabbit approximately 86 million years ago. Thus, in spite of some exceptional traits, the overall properties of the NMR genome appeared to be similar to those of other mammals.

Lineage-specific gene family expansions may be associated with the emergence of specific functions and physiology. Compared to other mammals, the NMR showed a moderate number of gene families under expansion and contraction (Fig. 1b), including 96 NMR lineage-specific gene families (Fig. 2). Analysis of syntenic regions identified 750 gained and 320 lost NMR genes (Supplementary Tables 12–14). At least 75.5% of genes gained showed evidence of transcription, and the lost genes were enriched for ribosome and nucleoside biosynthesis functions (Supplementary Table 15). We also identified 244 pseudogenes, containing 183 frameshift and 119 premature termination events (Supplementary Tables 16 and 17). Functional categories enriched for pseudogenes included olfactory receptor activity (GO:0004984, $P < 0.001$, Fisher's exact test, 36 genes), visual perception (GO:0007601, $P = 0.015$, *CRB1*, *CRYBB3*, *GNAT2*, *GRK7*, *GUCA1B* and *PDE6H*), spermatogenesis

Table 1 | Global statistics of the NMR genome

Sequencing	Insert size (bp)	Total data (Gb)	Sequence coverage (fold)
Paired end libraries	170–800	126.52	47
	2–20 × 10 ³	120.66	45
	Total	247.18	92
Assembly	N50 (kb)	Longest (kb)	Size (Gb)
Contigs	19.3	179	2.45
Scaffolds	1,585	7,787	2.66
Annotation	Number	Total length (Mb)	Percentage of the genome
Repeats	3,090,116	666.7	25
Genes	22,561	722.3	27.1
CDS	181,641	32.5	1.2

¹Department of Bioinspired Science, Ewha Womans University, Seoul, 120-750, Korea. ²BGI-Shenzhen, Shenzhen, 518083, China. ³Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁵Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁶Rodent Histopathology Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Department of Physiology and The Sam and Ann Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, Texas 78245, USA. ⁸Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA. ⁹Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark. ¹⁰Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark.

*These authors contributed equally to this work.

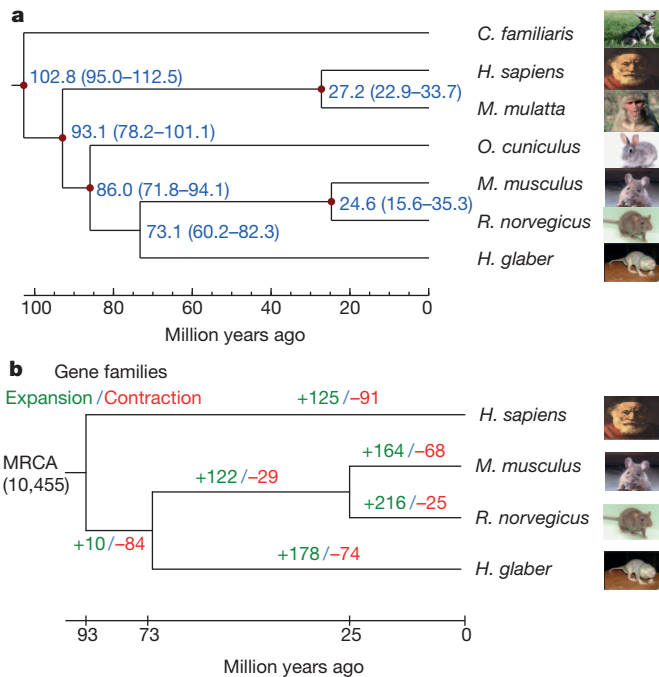


Figure 1 | Relationship of the NMR to other mammals. **a**, Estimation of the time of divergence (with error range shown in parentheses) of the NMR and six other mammals based on orthology relationship. Distances are shown in millions of years. **b**, Expansion and contraction in gene families. Numbers designate the number of gene families that have expanded (green) and contracted (red) since the split from the common ancestor. The most recent common ancestor (MCRA) has 10,455 gene families.

(GO:0007283, $P = 0.044$, *ADAM29*, *ADC*, *CCIN*, *CCT6B*, *DEDD*, *OAZ3* and *SHBG*), and possibly RING domain (SM00184, $P = 0.142$, *CNOT4*, *KCNRG*, *RNF5*, *TRIM17*, *TRIM11* and *ZSWIM2*). The enrichment in the visual perception category appears to underlie the evolution of poor vision in the NMR, whereas many RING-domain-containing proteins act as ubiquitin ligases¹². The levels of ubiquitinated proteins in NMRs are lower than in mice and, unlike those in mice, do not change significantly with age⁶.

Identification of genes that have undergone positive selection in the NMR lineage can provide useful pointers to the evolution of its unique

traits. 45 genes (0.4%) were identified as positively selected in the NMR lineage at the false discovery rate of 0.01 and 141 genes (1.2%) at the false discovery rate of 0.05 (Supplementary Table 18). 12 out of the 45 genes (corresponding to the false discovery rate of 0.01) passed a strict manual inspection for alignment quality. In comparison, 0.7% of genes were predicted to be positively selected in the human lineage from high-quality alignments and using Rom correction for multiple testing¹³. Interestingly, our set included *TEP1*, encoding a telomerase component, and *TERF1*, a telomeric repeat binding factor identified at the false discovery rate of 0.05 (Supplementary Fig. 10). The *TERF1* gene product is one of six proteins contributing to the shelterin complex, which shapes and protects telomeres¹⁴ and has been proposed to regulate telomere length¹⁵.

To gain further insights into biological processes that underlie the exceptional traits of the NMR, we identified 39 NMR proteins containing 45 amino acid residues unique among orthologues present in 36 vertebrate genomes (Supplementary Table 19). This gene set included cyclin E1 (*CCNE1*), uncoupling protein 1 (*UCP1*) and γ -crystallin (*CRYGS*), which are associated with the G1/S transition during the cell cycle, thermogenesis and visual function, respectively. Other noteworthy genes were *APEX1*, a multifunctional DNA repair enzyme, *RFC1*, replication factor C, and *TOP2A*, a DNA topoisomerase that controls the topologic states of DNA during transcription. This set also contained eight genes designated as cancer-related¹⁶. Finally, *TOP2A*, along with *TEP1* and *TERF1* from the set of positively selected genes, are part of a five-protein complex of alternate lengthening of telomere pathway¹⁷. Overall, these analyses point to altered telomerase function in the NMR, which may be related to its evolution of extended lifespan and cancer resistance.

We also identified 1.87 million heterozygous single-nucleotide polymorphisms (SNPs). This results in an estimated nucleotide diversity (mean per nucleotide heterozygosity) of 7×10^{-4} , which is much lower than in mouse and rat populations and is comparable to the nucleotide diversity observed in humans. Transition nucleotide changes were observed twice as often as transversions, indicating that variant calls reproduce the expected properties of natural variation in other mammals. This low level of nucleotide diversity may reflect a low effective size of NMR population, but may also be due to a high level of inbreeding, a reduced mutation rate or high efficiency of the repair systems. The variation of diversity along the genome was consistent with inbreeding in the NMR population. In protein-coding regions of the genome, our analysis identified 10,951 non-synonymous and 8,616 synonymous SNPs. Their ratio is much higher than in other studied organisms, including human, which appears to signal relaxation of purifying selection in the NMR, potentially as a consequence of reduced effective population size. Finally, we analysed the context dependency of NMR SNPs (Supplementary Fig. 11). Relative rates of nucleotide changes and nucleotide context dependencies were similar to those observed in human polymorphism, with the exception of a relative reduction of SNPs due to CpG mutations. This was caused by a combination of the relatively low CpG density in the NMR genome and a higher fraction of CpG dinucleotides within CpG islands compared to the human genome. CpG density was only 0.19 of that expected on the basis of the GC content, which is lower than in human, dog and panda genomes, but is similar to the mouse genome. However, in comparison to mouse, a higher fraction of CpG dinucleotides was concentrated in CpG islands. CpG dinucleotides within CpG islands contribute less to genetic variation because of their lower methylation rate and possibly also due to selection.

Long lifespan is a key feature of the NMR. To study ageing and longevity, we obtained RNA-seq data for brain, liver and kidney of newborn, young adult (4-year-old) and old adult (20-year-old) NMRs (Supplementary Table 20). In contrast to other mammals, few genes showed differential expression between 4- and 20-year-old NMRs, especially in the brain (Supplementary Tables 21–23). A recent study identified 33 underexpressed and 21 overexpressed genes in the human brain during ageing¹⁸. Of these, 32 genes did not show

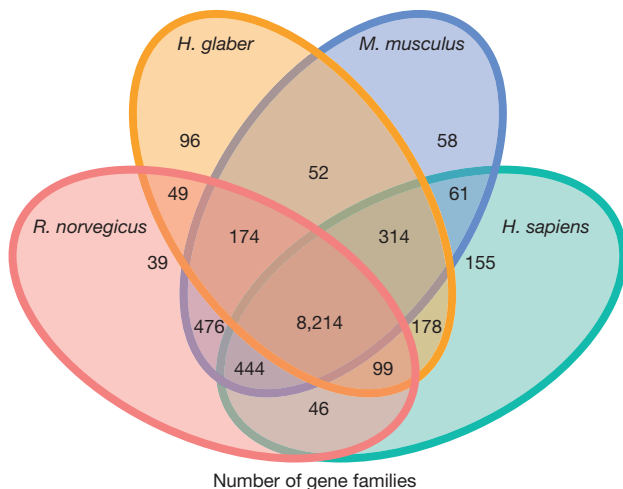


Figure 2 | Common and unique NMR gene families. This Venn diagram shows unique and overlapping gene families in the NMR (*H. glaber*), rat (*R. norvegicus*), mouse (*M. musculus*) and human (*H. sapiens*).

consistent expression changes with ageing in NMRs, including 30 genes that had stable expression and two genes that changed in the opposite direction compared to human brain (Supplementary Table 21). For example, *CYP46A1* and *SMAD3* were downregulated in the human brain, but showed elevated expression in the NMR brain. The product of the *CYP46A1* gene is a mediator of cholesterol homeostasis that influences the tendency of A β to aggregate. The product of *SMAD3* is a modulator of TGF- β signalling, playing a role in cancer development by slowing down the rate of cell proliferation. Elevated expression of *SMAD3* in the NMR during ageing may help optimize the rate of cell death, protecting NMRs from cancer.

A previous meta-analysis of age-related gene expression in mice, rats and humans revealed 56 consistently overexpressed and 17 under-expressed genes¹⁹. However, many of these genes did not show the same expression changes, suggesting that different regulatory mechanisms may underlie NMR longevity (Supplementary Tables 22 and 23). For example, genes related to degradation of macromolecules, such as *GSTA1*, *DERL1* and *GNS*, were not upregulated with age in NMRs. We also found that genes encoding mitochondrial proteins (*NDUFB11*, *ATP5G3* and *UQCRCQ*) were not downregulated, consistent with stable maintenance of mitochondrial function during ageing. It is also of interest that *TERT* (telomerase reverse transcriptase) showed stable expression regardless of age (Supplementary Fig. 12). This finding is consistent with the role of the telomerase complex, highlighted by positive selection on *TEP1* and *TERF1*. Overall, transcriptome and sequence data revealed different (compared to humans, mice and rats) patterns of NMR genes, which may underlie longevity mechanisms in this animal.

Non-shivering thermogenesis is a major heat production process in mammals that mainly depends on the action of *UCP1*, one of the 39 vertebrate genes that changed uniquely in the NMR (Supplementary Table 19). *UCP1* featured changes in amino acids Gln 146, Arg 263, Trp 264 and Thr 303, with the latter two residues being subject to positive selection ($P < 0.05$, likelihood ratio test for the branch-site model, $n = 30$) and Arg 263 and Trp 264 located in the conserved nucleotide binding motif (Fig. 3a). With Arg–Trp instead of the rigid Gly–Pro in the key regulatory site, *UCP1* is expected to lose the tight regulation by purine nucleotides as inhibitors and fatty acids as activators (Fig. 3b and c). The same loop also features two positively

charged Lys residues followed by a negatively charged residue (also a unique combination), that should markedly affect the local electrostatic potential of *UCP1*. In addition, Gln 146 replaced a conserved His involved in proton transport, and the same mutation was shown to decrease proton conductance of *UCP1* fivefold²⁰. Thr 303 is located in the carboxy-terminal motif (RqTxDCxT) required for binding purine nucleotides²¹. Taken together, these observations indicate a tight association of *UCP1* function with the unique thermoregulation of the NMR²².

In mammals, switches between light and dark periods affect synthesis of the hormone melatonin, which modulates sleep and circadian rhythms. NMRs live in a naturally dark habitat and their pineal glands, where melatonin is synthesized, are atrophied²³, but we found that the genes involved in melatonin synthesis (*TPH1*, *TPH2*, *DDC*, *AANAT* and *ASMT*) are intact. Interestingly, the expression of genes involved in the final two steps of melatonin synthesis was very low (*AANAT*) or undetectable (*ASMT*) in the NMR brain regardless of age (Supplementary Table 24 and Supplementary Fig. 13). Moreover, two major mammalian melatonin receptors (*MTNRIA* and *MTNR1B*, encoding MT₁ and MT₂, respectively) were inactivated by mutations that introduce premature stop signals (Supplementary Fig. 14). Synteny analyses showed that these pseudogenes corresponded to mouse *MTNRIA* and *MTNR1B*. Although melatonin signalling appears to be disrupted in the NMR, its circadian rhythms were maintained in terms of locomotor activity and body temperature when exposed to periodic light/dark changes²⁴. Our finding is consistent with a previous report that MT₁/MT₂ knockout mice maintained essentially normal circadian rhythms²⁵. These mice also showed decreased insulin secretion²⁵. Likewise, our transcriptome analysis of the NMR revealed decreased expression of genes involved in insulin/IGF-1 signalling in the liver compared to mice (Supplementary Fig. 15).

To explain the extraordinary resistance of the NMR to cancer³, a two-tier protective mechanism involving contact inhibition mediated by p16^{Ink4a} and p27^{Kip1} was proposed⁴. The involvement of p16^{Ink4a} is unusual, since humans and mice show only contact inhibition mediated by p27^{Kip1}. We analysed the gene locus and the transcriptome reads corresponding to tumour suppressors p16^{Ink4a} and p19^{Arf}. As in mice, the p16^{Ink4a} transcript consists of three exons (Supplementary Fig. 16). However, sequence similarity in the last exon is low, and two early stop codons in the second exon were predicted to result in a shorter, 14-kDa protein (Supplementary Fig. 17). The four ankyrin repeats were, however, intact and Thr69, a residue important for CDK6 binding, was conserved, so the function of the protein may be partially preserved (Supplementary Fig. 18). The p19^{Arf} transcript consists of two exons, but four stop codons in the second exon should lead to a shorter, 10-kDa protein (Supplementary Figs 19–21).

The NMR is also unique in that its skin and cutaneous C-fibres lack the neuropeptide Substance P, making this animal insensitive to certain types of pain^{10,11}. Our analysis revealed the presence of intact *TAC1* encoding Substance P. However, the NMR had a deletion in the core promoter region highly conserved among mammals (Supplementary Fig. 22). Thus, this neurotransmitter appears to be functional but may be under unique regulation.

We further examined the molecular basis for poor visual function and small eyes in the NMR. Of the four vertebrate opsin genes (*RHO*, *OPN1LW*, *OPN1MW* and *OPN1SW*), two (*OPN1LW* and *OPN1MW*) were missing (Table 2); this distinguishes the NMR from other rodents with dichromatic colour vision, such as mice, rats and guinea pigs. However, the NMR has intact *RHO* (rhodopsin) and *OPN4* (melanopsin), supporting the presence of rod-dominated retinæ and the capacity to distinguish light/dark cues. Of about 200 genes associated with visual perception (GO:0007601) in humans and mice, almost 10% were inactivated or missing in the NMR (Table 2 and Supplementary Fig. 23). These mammalian genes participate in crystallin formation, phototransduction in the retina, retinal development, dark adaptation, night blindness and colour vision. For at least ten of these

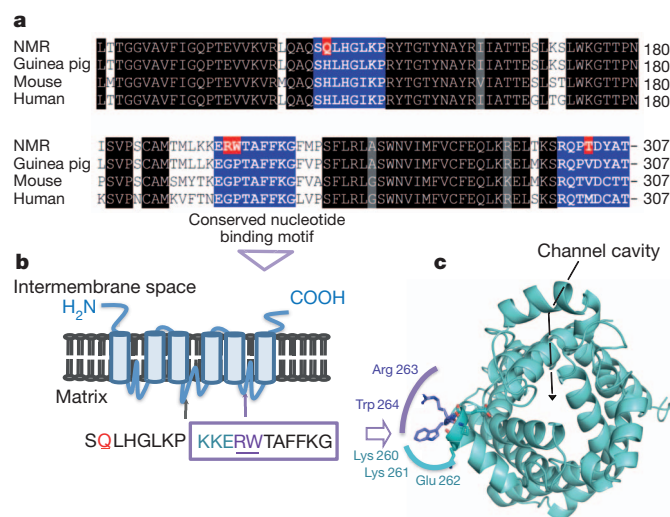


Figure 3 | Unique changes in UCP1 sequences and their roles in thermoregulation. **a**, Alignment of mammalian UCP1 sequences. Amino acids unique to the NMR are highlighted in red, and conserved motifs in blue. **b**, Topology of UCP1. Regions affected in the NMR are highlighted. **c**, Structural model of UCP1. Location of the channel and the nucleotide-binding loop with altered sequences in the NMR are shown.

Table 2 | Visual perception genes that are inactivated or are missing in the NMR genome

Gene	Inactivation event	Time of gene loss	ω_0 (average)	ω_1 (other)	ω_2 (NMR)	P-value
<i>RBP3</i>	F	NMR	0.121	0.085	0.232	1.187E-11
<i>ARR3</i>	F/S	NMR	0.420	0.260	0.912	6.263E-06
<i>PDE6C</i>	F	NMR	0.171	0.139	0.316	0.0001
<i>GUCA1B</i>	F	NMR	0.083	0.056	0.217	0.001
<i>GJA10</i>	F/S	NMR	0.308	0.248	0.524	0.002
<i>GUCY2E</i>	F	NMR	0.124	0.105	0.182	0.002
<i>CRYBA4</i>	S	NMR	0.055	0.036	0.123	0.001
<i>GNAT2</i>	F	NMR	0.055	0.039	0.108	0.017
<i>SLC24A1</i>	F	NMR	0.389	0.355	0.517	0.035
<i>CRYBB3</i>	S	NMR	0.071	0.048	0.122	0.037
<i>RP1L1</i>	S	NMR	0.448	0.424	0.513	0.186
<i>GRK7</i>	F	NMR	0.154	0.135	0.201	0.335
<i>PDE6H</i>	F	NMR	0.091	0.082	0.127	0.648
<i>EYS</i>	F/S	Ancestor	-	-	-	-
<i>GUCA1C</i>	F	Ancestor	-	-	-	-
<i>OPN1LW</i>	L	NMR	-	-	-	-
<i>OPN1MW</i>	L	NMR	-	-	-	-
<i>PRCD</i>	L	NMR	-	-	-	-
<i>RD3</i>	L	NMR	-	-	-	-

F, frameshift; S, premature stop codon; L, complete loss detected with synteny information. Ancestor/NMR indicates that the gene was lost in a rodent ancestor or the NMR, respectively. The rate ratio ω of non-synonymous to synonymous substitutions was calculated by using human, mouse, rat and NMR sequences. ω_0 is the average ratio in all branches, ω_1 is the average ratio in non-NMR branches, and ω_2 is the ratio in the NMR branch. A small P-value indicates that the two-ratio model fits the data better than the one-ratio model.

genes, we observed relaxation of the functional constrain on NMR sequences by estimating the ratio of non-synonymous to synonymous substitutions, which corroborated the dysfunction of these genes. Inactivation of *CRYBA4*, a microphthalmia-related gene, may be associated with the small-sized eyes, whereas inactivation of *CRYBA4* and *CRYBB3* and a NMR-specific mutation in *CRYGS* (Supplementary Table 19) may be associated with abnormal eye morphology²⁶. Thus, while some genes responsible for vision are preserved in the NMR, its poor visual function may be explained by deterioration of genes coding for various critical components of the visual system.

Further analysis revealed substantial divergence of the NMR nuclear receptor corepressor Hairless from other mammalian orthologues and the presence of amino acid replacements associated with the hairless phenotype, which is consistent with the lack of fur in NMRs (Supplementary Fig. 24). In addition, we found substantial sequence variation in the sweet taste receptor and lack of many bitter taste receptors common to other mammals (Supplementary Fig. 25 and 26). In particular, the NMR appears to lack the phenylthiocarbamide taste, a dominant genetic trait in humans, as well as several other common bitter tastes.

Air in NMR burrows is low in O₂ (~8%) and high in CO₂ (>10%) owing to many animals sharing a limited air supply and poor gas exchange through soil²⁷. To cope with the low O₂ conditions, the NMR has developed adaptive circulatory (altered haemoglobin oxygen affinity) and metabolic functions, reducing metabolic rate and slowing down development^{1,8,28,29}. To obtain insights into this adaptation, we examined gene expression changes in several tissues of NMR subjected to 8% O₂ for one week (Supplementary Tables 25–31 and Supplementary Fig. 27–30). Many changes associated with energy metabolism and redox control were observed. Sequence analysis of NMR hypoxia-induced factor 1 α (HIF1 α) revealed a T407I exchange unique among mammals and located in the VHL-binding domain. Under normal oxygen conditions, VHL mediates ubiquitin-dependent degradation of HIF1 α . In addition, NMR VHL harbours V166I exchange at a functionally important site. These amino acid changes are consistent with relaxation of ubiquitin-dependent degradation of HIF1 α , and, thus, with adaptation to low oxygen conditions.

To summarize, sequencing and analysis of the NMR genome revealed numerous insights into the biology of this remarkable animal. In addition, this genome and the associated data sets offer the research

communities working in ageing, cancer, eusociality and many other areas a rich resource that can be mined in numerous ways to uncover the molecular bases for the extraordinary traits of this most unusual mammal. In turn, this information provides unprecedented opportunities for addressing some of the most challenging questions in biology and medicine, such as mechanisms of ageing, the role of genetic makeup in regulating lifespan, adaptations to extreme environments, hypoxia tolerance, thermogenesis, resistance to cancer, circadian rhythms, sexual development and hormonal regulation.

METHODS SUMMARY

The NMR genome was sequenced on the Illumina HiSeq 2000 platform. The sequenced individual male NMR was from a captive breeding colony located at the University of Illinois, Chicago. The genome was assembled using SOAPdenovo. We obtained 2.5 Gb (gigabase pairs) contig sequences with N50 19.3 kb (kilobase pairs) and N90 4.7 kb, and 2.7 Gb scaffold sequences with N50 1.6 Mb (megabase pairs) and N90 0.3 Mb. (The N50 (or N90) contig size is the length of the smallest contig S in the sorted list of all contigs where the cumulative length from the largest contig to contig S is at least 50% (or 90%) of the total assembly length.) RNA-seq data (ageing and low O₂ experiments) were for animals from the same colony. See Supplementary Information for data analysis and additional details.

Received 14 June; accepted 5 September 2011.

Published online 12 October 2011.

- Edrey, Y. H., Park, T. J., Kang, H., Biney, A. & Buffenstein, R. Endocrine function and neurobiology of the longest-living rodent, the naked mole-rat. *Exp. Gerontol.* **46**, 116–123 (2011).
- Buffenstein, R. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *J. Comp. Physiol. B* **178**, 439–445 (2008).
- Liang, S., Mele, J., Wu, Y., Buffenstein, R. & Hornsby, P. J. Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (*Heterocephalus glaber*). *Aging Cell* **9**, 626–635 (2010).
- Seluanov, A. et al. Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proc. Natl Acad. Sci. USA* **106**, 19352–19357 (2009).
- Andziak, B. et al. High oxidative damage levels in the longest-living rodent, the naked mole-rat. *Aging Cell* **5**, 463–471 (2006).
- Pérez, V. I. et al. Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl Acad. Sci. USA* **106**, 3059–3064 (2009).
- Jarvis, J. U. M. Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science* **212**, 571–573 (1981).
- Larson, J. & Park, T. J. Extreme hypoxia tolerance of naked mole-rat brain. *Neuroreport* **20**, 1634–1637 (2009).
- Buffenstein, R., Woodley, R., Thomadakis, C., Daly, T. J. & Gray, D. A. Cold-induced changes in thyroid function in a poikilothermic mammal, the naked mole-rat. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **280**, R149–R155 (2001).
- Park, T. J. et al. Selective inflammatory pain insensitivity in the African naked mole-rat (*Heterocephalus glaber*). *PLoS Biol.* **6**, e13 (2008).
- Smith, E. S., Blass, G. R., Lewin, G. R. & Park, T. J. Absence of histamine-induced itch in the African naked mole-rat and “rescue” by Substance P. *Mol. Pain* **6**, 29 (2010).
- Deshaies, R. J. & Joazeiro, C. A. P. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
- Schneider, A. et al. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol. Evol.* **1**, 114–118 (2009).
- de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
- van Steensel, B. & Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743 (1997).
- Higgins, M. E. et al. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.* **35**, D721–D726 (2007).
- Bhattacharyya, S. et al. Telomerase-associated protein 1, HSP90, and topoisomerase IIA associate directly with the BLM helicase in immortalized cells using ALT and modulate its helicase activity using telomeric DNA substrates. *J. Biol. Chem.* **284**, 14966–14977 (2009).
- Hong, M. G., Myers, A. J., Magnusson, P. K. E. & Prince, J. A. Transcriptome-wide assessment of human brain and lymphocyte senescence. *PLoS ONE* **3**, e3024 (2008).
- de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of ageing. *Bioinformatics* **25**, 875–881 (2009).
- Bienengraeber, M., Echta, K. S. & Klingenberg, M. H⁺ transport by uncoupling protein (UCP-1) is dependent on a histidine pair, absent in UCP-2 and UCP-3. *Biochemistry* **37**, 3–8 (1998).
- Porter, R. K. Uncoupling protein 1: a short-circuit in the chemiosmotic process. *J. Bioenerg. Biomembr.* **40**, 457–461 (2008).

22. Woodley, R. & Buffenstein, R. Thermogenic changes with chronic cold exposure in the naked mole-rat (*Heterocephalus glaber*). *Comp. Biochem. Physiol. A* **133**, 827–834 (2002).
23. Quay, W. B. Pineal atrophy and other neuroendocrine and circumventricular features of the naked mole-rat, *Heterocephalus glaber* (Rüppell), a fossorial, equatorial rodent. *J. Neural Transm.* **52**, 107–115 (1981).
24. Riccio, A. P. & Goldman, B. D. Circadian rhythms of locomotor activity in naked mole-rats (*Heterocephalus glaber*). *Physiol. Behav.* **71**, 1–13 (2000).
25. Mühlbauer, E., Gross, E., Labucay, K., Wolgast, S. & Peschke, E. Loss of melatonin signalling and its impact on circadian rhythms in mouse organs regulating blood glucose. *Eur. J. Pharmacol.* **606**, 61–71 (2009).
26. Nikitina, N. V. et al. Postnatal development of the eye in the naked mole rat (*Heterocephalus glaber*). *Anat. Rec. A* **277A**, 317–337 (2004).
27. Bennett, N. C. & Faulkes, C. G. *African Mole-Rats: Ecology and Eusociality* (Cambridge University Press, 2000).
28. McNab, B. K. The influence of body size on the energetics and distribution of fossorial and burrowing mammals. *Ecology* **60**, 1010–1021 (1979).
29. Johansen, K., Lykkeboe, G., Weber, R. E. & Maloiy, G. M. Blood respiratory properties in the naked mole rat *Heterocephalus glaber*, a mammal of low body temperature. *Respir. Physiol.* **28**, 303–314 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. C. LaVinka for help in preparing NMR tissues and N. Estolaga for help in preparing the manuscript. We acknowledge financial support from the WCU Program (R31-2008-000-10010-0), the NIH (AG038004, AG021518 and CA080946), the Shenzhen Municipal Government (ZYC200903240077A), the

National Natural Science Foundation of China (30725008) and the National Science Foundation (0744979).

Author Contributions V.N.G. conceived the study. T.J.P. carried out animal work. A.A.T., M.V.K. and S.H.Y. prepared samples. X.F., Z.H., L.H., X.S., P.Y., X.Z., C.P., Z.X., Y.Z., Y.C., Q.Z., L.Y., B.W., C.H., Q.L., L.C., W.Z., G.Z. and J.W. performed genome sequencing and assembly. X.F., G.Z. and J.W. supervised genome sequencing and assembly. E.B.K., X.F., A.A.F., Z.H., A.V.L., S.M.M., L.P., G.Z. and V.N.G. performed genome and transcriptome analyses. Z.H., N.S., P.P., A.K. and S.R.S. carried out genetic analyses. G.V.K., R.T.B. and R.B. discussed the data. All authors contributed to data interpretation. V.N.G. wrote the paper with significant contributions from E.B.K., X.F., A.A.F., Z.H., S.R.S. and G.Z., and input from all authors.

Author Information The NMR whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AFSB00000000. The version described in this paper is the first version, AFSB01000000. The mitochondrial sequence has been deposited at GenBank under the accession number JN242813. All short-read data have been deposited into the Short Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under the accession number SRA030468. Raw sequencing data of the transcriptome have been deposited in the Gene Expression Omnibus with the accession number GSE30337. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.Z. (zhanggj@genomics.org.cn), J.W. (wangj@genomics.org.cn) or V.N.G. (vgladyshev@rics.bwh.harvard.edu).

Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase

Mardo Kõivomägi¹, Ervin Valk¹, Rainis Venta¹, Anna Iofik¹, Martin Lepiku¹, Eva Rose M. Balog², Seth M. Rubin², David O. Morgan³ & Mart Loog¹

Multisite phosphorylation of proteins has been proposed to transform a graded protein kinase signal into an ultrasensitive switch-like response^{1–4}. Although many multiphosphorylated targets have been identified, the dynamics and sequence of individual phosphorylation events within the multisite phosphorylation process have never been thoroughly studied. In *Saccharomyces cerevisiae*, the initiation of S phase is thought to be governed by complexes of Cdk1 and Cln cyclins that phosphorylate six or more sites on the Clb5–Cdk1 inhibitor Sic1, directing it to SCF-mediated destruction^{1,5–8}. The resulting Sic1-free Clb5–Cdk1 complex triggers S phase⁹. Here, we demonstrate that Sic1 destruction depends on a more complex process in which both Cln2–Cdk1 and Clb5–Cdk1 act in processive multiphosphorylation cascades leading to the phosphorylation of a small number of specific phosphodegrons. The routes of these phosphorylation cascades are shaped by precisely oriented docking interactions mediated by cyclin-specific docking motifs in Sic1 and by Cks1, the phospho-adaptor subunit of Cdk1. Our results indicate that Clb5–Cdk1-dependent phosphorylation generates positive feedback that is required for switch-like Sic1 destruction. Our evidence for a docking network within clusters of phosphorylation sites uncovers a new level of complexity in Cdk1-dependent regulation of cell cycle transitions, and has general implications for the regulation of cellular processes by multisite phosphorylation.

To study the multiphosphorylation of Sic1, we used a non-inhibitory truncated version of Sic1 (Sic1ΔC) as a substrate for purified Cln2–Cdk1 and Clb5–Cdk1 complexes (Supplementary Fig. 1a; Cdk1 is also known as Cdc28). Intriguingly, both Cln2–Cdk1 and Clb5–Cdk1 generated phosphorylation patterns with abruptly accumulating hyperphosphorylated species (Fig. 1a, b, d, e and Supplementary Fig. 1b). This pattern depended on Cks1, the Cdk1 subunit that binds phosphate groups¹⁰. Mutation of the phosphate-binding site of Cks1 reduced the accumulation of multiphosphorylated forms (Fig. 1a and Supplementary Fig. 2a). Similarly, a phosphorylated competitor phosphopeptide reduced phosphorylation (Fig. 1b). Cks1 mutation and the peptide had little effect on the phosphorylation of a Sic1ΔC version containing a single Cdk site (Fig. 1c).

Cks1 is essential for Cdk1 function^{11,12}, with roles at the G₁/S and G₂/M transitions^{13,14}. We found that the Cks1:Cdk1 stoichiometry *in vivo* was about 1:1 for Cln2–Cdk1 and at least 0.5:1 for Clb5–Cdk1, confirming that Cks1-dependent multiphosphorylation is the prevalent mode of Cdk1 action *in vivo* (Supplementary Fig. 1a–c). An isothermal calorimetry binding assay of fully phosphorylated Sic1ΔC (pSic1ΔC) and Cks1 measured a K_D of $11 \pm 2 \mu\text{M}$, whereas the non-phosphorylated version showed no detectable binding (Supplementary Fig. 1d). Approximately three to four molecules of Cks1 bound each molecule of pSic1ΔC, suggesting that several phosphorylated sites can bind Cks1 independently. Finally, we found that the phospho-binding pocket of Cks1 is required for phosphorylation and degradation of Sic1 *in vivo* (Supplementary Fig. 1e, f).

To understand the Cks1-dependent mechanism, we analysed Sic1ΔC multiphosphorylation over time (Fig. 1d, e). We did not observe significant accumulation of intermediate phosphorylated forms, indicating that phosphorylation was processive. When we performed kinase reactions at Sic1ΔC concentrations higher than apparent K_M , multiphosphorylation patterns remained constant despite the increase in the inhibition term $1 + [S_{OP}]/K_M$ (where S_{OP} is unphosphorylated substrate) (Supplementary Fig. 2b–e). Thus, the enzyme displays processivity; that is, it is able to transfer two or more phosphates to the substrate during a single association event. This conclusion was additionally confirmed using different enzyme concentrations

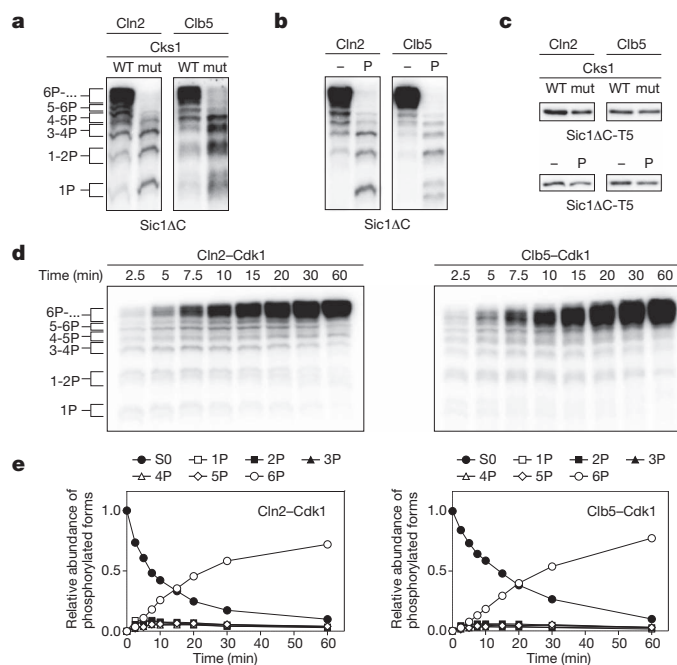


Figure 1 | The phospho-adaptor subunit Cks1 provides processivity for the multiphosphorylation of Sic1 by Cln2–Cdk1 and Clb5–Cdk1. **a**, Cln2–Cdk1 and Clb5–Cdk1 complexes were incubated with Sic1ΔC and [γ -³²P]ATP. The reactions also included wild-type Cks1 (WT) or a version with a mutated phosphate-binding site (mut; see Supplementary Methods). Phosphorylated substrates were separated using Phos-tag SDS–PAGE gels. **b**, Reactions were performed in the presence of a phosphopeptide competitor (P) based on the sequence surrounding T45 in Sic1. **c**, The phosphorylation of a Sic1ΔC version containing a single Cdk site (Sic1ΔC-T5, with other Cdk consensus sites mutated to alanines) was not affected by the Cks1 mutant (mut) or the phosphopeptide. Standard SDS–PAGE was used. **d**, Time courses of Sic1ΔC multiphosphorylation were followed by Phos-tag SDS–PAGE. **e**, The quantified data from **d**. The intensities of ³²P-labelled proteins were divided by the number of phosphates as indicated to obtain the levels of different phosphoforms (see also Methods). In the experiments presented in Fig. 1, enzyme concentrations were chosen to obtain roughly equal substrate labelling.

¹Institute of Technology, University of Tartu, Tartu 50411, Estonia. ²Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064, USA. ³Department of Physiology, University of California, San Francisco, California 94158, USA.

in the assay (Supplementary Fig. 2f) and in mathematical simulations (Supplementary Fig. 3). This processive pattern argues against the current model of ultrasensitivity in the Sic1 phosphorylation switch, which is based on the assumption of a distributive mechanism with equal specificity of different sites^{1,15}.

To dissect the mechanism of the processive multiphosphorylation cascade, we first studied the impact of potential docking interactions between Sic1 and cyclins. In previous studies, we found that rapid Sic1 phosphorylation by Clb5–Cdk1 depends on an interaction between RXL motifs in Sic1 and the hydrophobic patch docking site (hp) in Clb5; a triple mutation in this site (Clb5hpm) decreases the net phosphorylation rate (Supplementary Fig. 4a–h)¹⁶. Further, we found here that a version of Sic1ΔC with mutations at its four RXL motifs (Sic1ΔC-1234rxl) showed less abrupt production of multiphosphorylated species by Clb5–Cdk1, showing that processive multiphosphorylation requires both Cks1-dependent and hp-dependent docking (Supplementary Fig. 4i). Cln2–Cdk1 had only a mild RXL effect on the phosphorylation pattern, probably because Cln2 does not contain a conventional hp like that in the B-type cyclins. In recent studies, we also located a ten-amino-acid stretch in Sic1, 136VLLPPSRPT145, which confers Cln2 specificity¹⁶. Here we found that a five-alanine mutation of the first five hydrophobic residues in this stretch, or a synthetic competitor peptide containing the docking site, reduced the abrupt multiphosphorylation pattern for Cln2 (Supplementary Fig. 4j). In conclusion, both Clb5–Cdk1 and Cln2–Cdk1 use docking mechanisms, in addition to Cks1, to achieve processive multiphosphorylation of Sic1.

Using Sic1ΔC mutants carrying only one Cdk site (Fig. 2a, b), we found that Clb5–Cdk1 rapidly phosphorylated just four sites (T5, T33,

S76 and S80), and this specificity depended on the RXL2 and RXL3 docking sites in Sic1 (Fig. 2b; note that in all figure labels, the indicated Cdk sites are those left unmutated, unless otherwise indicated). Cln2–Cdk1, on the other hand, showed a preference for the N-terminally located site T5 (Fig. 2b). Thus, docking interactions direct the associated kinase to a small number of primary phosphorylation sites. We speculate that these primary sites interact with Cks1 to drive processive phosphorylation of additional sites.

With these primary specificities in mind, we set out to map the pathways along which Cln2–Cdk1 and Clb5–Cdk1 catalyse the phosphorylation of the critical sites required for Sic1 degradation. The original model of Sic1 regulation proposed that six or more sites must be simultaneously phosphorylated *in vivo* to facilitate binding of phospho-Sic1 to the SCF subunit Cdc4 (ref. 1). On the other hand, later binding studies revealed that closely positioned pairs of phosphorylation sites (pT5/pS9, pT45/pT48, or pS76/pS80; see Fig. 2a) each present separate entities with a strong affinity for Cdc4, indicating that just two phosphorylation sites, in the right positions, might be sufficient for Sic1 degradation¹⁷. Our results provided a way to reconcile these findings: we proposed that the requirement for six or more sites *in vivo* reflects a requirement for priming phosphorylation events that direct processive phosphorylation of critical phosphodegrons. To test this possibility, we first measured phosphorylation of a Sic1ΔC mutant with all Cdk sites changed to alanine except for the triple cluster S69/S76/S80, which contains two potential paired degrons (S69/S76 and S76/S80). There was no processive multiphosphorylation of the cluster S69/S76/S80 (Fig. 2c, lane 2 in each panel), but processivity could be induced by adding back single Cdk1 sites to

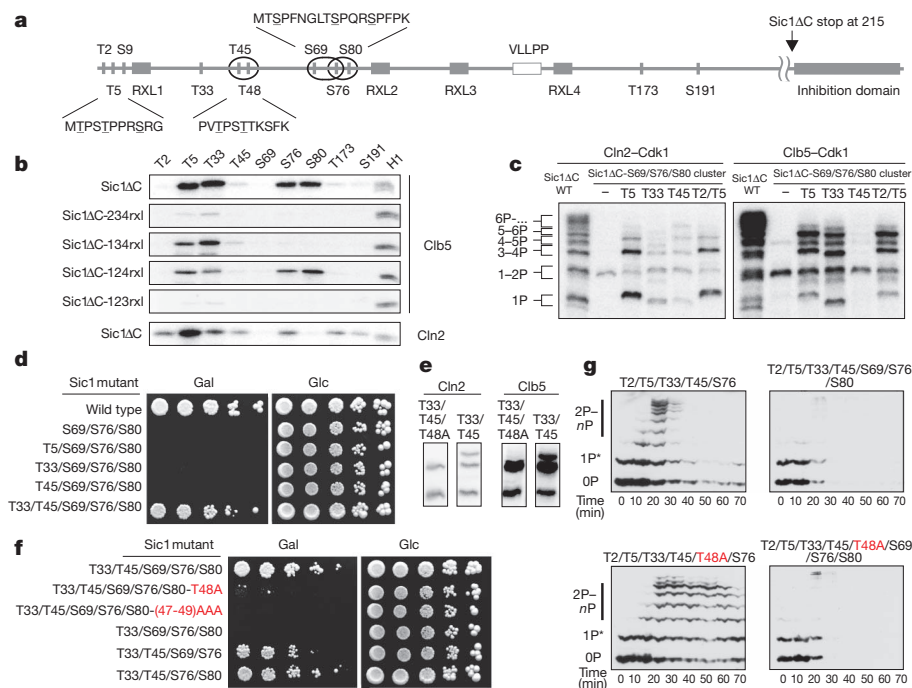


Figure 2 | Phosphorylated priming sites provide docking interactions for efficient phosphorylation of suboptimal sites in phosphodegrons.

a, Schematic view of phosphorylation sites, docking motifs (Clb5- and Cln2-specific), phosphodegrons (ovals¹⁷) in Sic1. **b**, Phosphorylation specificity of Clb5–Cdk1 and Cln2–Cdk1 towards different Cdk sites was studied using Sic1ΔC constructs containing a single fixed Cdk site. For Clb5–Cdk1, the dependence of the site-specificity profile on RXL docking sites was assessed using Sic1ΔC constructs containing a single Cdk site and a single fixed RXL motif. **c**, The impact of different priming phosphorylation sites on cooperative phosphorylation of the degron cluster S69/S76/S80. Phospho-site mutants of Sic1ΔC carrying the intact S69/S76/S80 cluster and the indicated sites left unmutated were used in a kinase assay with Cln2–Cdk1 and Clb5–Cdk1 using Phos-tag SDS–PAGE. **d**, Full-length Sic1 versions containing the combination

of sites described in **c** were overexpressed under the galactose promoter to assay the ability of cells to degrade Sic1. Gal, galactose; Glc, glucose. **e**, Comparison of the *in vitro* phosphorylation profiles of Sic1ΔC versions containing only the phosphorylation sites T33/T45 or T33/T45 with mutation T48A. **f**, The nonconsensus Cdk1 site T48 is important for viability of cells overexpressing Sic1. The same assay as **d** was used. In panels **d** and **f**, the labels indicate unmutated amino acids, and all other consensus Cdk sites are mutated; mutations in the nonconsensus Cdk sites are highlighted in red. **g**, The phosphorylation and degradation dynamics of Sic1 were followed after the release of cells from α -factor in a system constitutively expressing mutated versions of noninhibitory haemagglutinin-tagged Sic1ΔC–3HA. The asterisk indicates a G₁-specific phosphorylation by an unknown kinase.

the amino-terminal side of the cluster. The rate of the appearance of multiphosphorylated species correlated with the site specificity data for Cln2 and Clb5 (Fig. 2b). Addition of the most Cln2-specific site, T5, caused a much greater effect in Cln2 reactions than the less Cln2-specific sites, T33 and T45. Addition of the Clb5-specific sites T5 and T33, but not the poor Clb5 site T45, greatly increased processivity in Clb5 reactions. These results indicate that sites T5, T33 and T45 are able to serve as priming sites for Cks1-dependent phosphorylation of the S69/S76/S80 degreon cluster. The phosphopeptide-dependence was confirmed for these mutants as described for Sic1 Δ C-wild type in Fig. 1b (data not shown).

Cells overexpressing Sic1 containing only the triple cluster S69/S76/S80 were inviable (Fig. 2d). Addition of T5, T33 or T45 did not prevent this lethal effect, indicating that phosphorylation of the S69/S76/S80 cluster alone is not sufficient for degradation of Sic1. However, viability improved when both T33 and T45 were added. Notably, the addition of T45, together with a nonconsensus Cdk site T48 (Fig. 2a), creates a double degreon, T45/T48, as predicted previously¹⁷. T48 is known to be phosphorylated *in vivo*⁷. We proposed that T33 serves as a docking site for both the T45/T48 and S69/S76/S80 clusters, and that T45 serves both as a constituent site of the T45/T48 degreon as well as a Cks1-dependent docking site for the degreon cluster S69/S76/S80. Indeed, T48 alone (the Sic1-9A mutant with all Cdk sites changed to alanine) was a very poor substrate for Clb5 and Cln2 (data not shown), but the addition of T33 or the T33/T45 pair made it a specific site (Fig. 2e, Supplementary Fig. 5 and Supplementary Table 1), implying that T33 phosphorylation allows bypass of the consensus motif requirement of a +1 proline at T48. Notably, we found that the alanine mutation in T48 had a strong growth-suppressing effect in the galactose assay within the context of sites T33/T45/S69/S76/S80 and also had a weaker effect in the background containing all the Cdk sites (Fig. 2f and Supplementary Fig. 4k). Our results indicate that the benefit of multisite phosphorylation of Sic1, compared to a system with a single phosphorylated site with high affinity for Cdc4 (for example, based on an optimal degreon site of cyclin E¹), is likely to be the ability of this mechanism to provide

docking-dependent kinase specificity for paired degreons. Thus, phosphorylation sites in Sic1 can be divided into three categories: (1) paired degreon sites that are spaced 3–7 amino acids from each another; (2) N-terminally positioned priming sites for each paired degreon; and (3) sites that serve as both priming and degreon sites (for example, T45).

It is unclear why a single paired degreon with a priming site is insufficient for degradation of Sic1, as suggested by our observation that viability in the galactose plate assay requires both T45/T48 and either S69/S76 or S76/S80 (Fig. 2f). To test if the simultaneous presence of both paired degreons is required for degradation of Sic1, we used western blotting to measure phosphorylation and degradation of mutated versions of Sic1 Δ C. Remarkably, a Sic1 Δ C construct (equivalent to the Sic1-5p mutant of ref. 1) containing only the degreon T45/T48, but missing the intact degreon around S76, was rapidly degraded after the release of cells from G₁ (Fig. 2g). This degradation was abolished by mutation of the single nonconsensus Cdk site at T48. A construct also containing the S69/S76/S80 degreon (Sic1-7p of ref. 1) was more rapidly degraded, and mutation of T48 in this background did not influence the degradation rate. We conclude that the T45/T48 degreon is sufficient to promote some degradation of Sic1 *in vivo*, but this rate of degradation is not sufficient to prevent the lethality of overexpressed Sic1.

Our model assumes differential roles of Cln2 and Clb5 in the order of Sic1 phosphorylation events. To explore this possibility, we developed methods for determining the apparent rate constants, which we termed k_{dock} , for individual Cks1-enhanced phosphorylation steps (Fig. 3a, b and Supplementary Table 1). The results revealed considerable differences between Clb5 and Cln2. Clb5–Cdk1 was much more effective than Cln2–Cdk1 in taking shortcuts to the critical degreon pair of S76/S80, using T5 and T33 as priming sites for Cks1, and with assistance from RXL-mediated docking (Supplementary Tables 1 and 2, Fig. 3a–c and Supplementary Fig. 6). Notably, in the case of Clb5, different RXL motifs supported different Cks1-dependent docking events (Supplementary Table 2).

We propose that in late G₁, Clb5–Cdk1 is inhibited by Sic1, and the cascade of phosphorylation events begins with T5 phosphorylation by

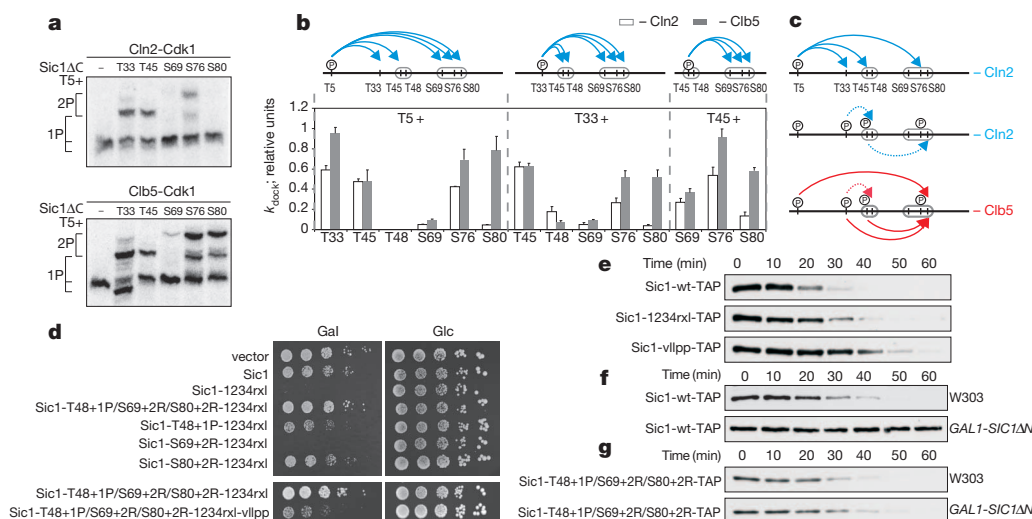


Figure 3 | Differential roles of Cln2 and Clb5 in Sic1 multiphosphorylation and degradation. **a**, Pair-wise mapping of the docking connections underlying Sic1 multiphosphorylation, using purified Sic1 Δ C mutants containing just two of the Cdk phosphorylation sites per mutant. Representative examples of autoradiographs of phosphorylation assays, showing different docking specificities between Cln2–Cdk1 and Clb5–Cdk1. **b**, The specificity profiles for different pair-wise docking connections. The error bars indicate standard errors of the means of at least two independent experiments. (See Supplementary Information and Supplementary Table 1). **c**, Schematic view of fast and slow docking-dependent phosphorylation steps for Cln2–Cdk1 and Clb5–Cdk1. **d**, Sic1 mutants with improved Cdk recognition determinants in suboptimal phosphodegreons rescue the inviability of cells overexpressing Sic1-

1234rxl. The Cln2-dependent docking site becomes essential under these conditions. **e**, Cells carrying *SIC1wt-TAP* or versions with docking site mutations at the endogenous *SIC1* locus were released from an α -factor arrest and the degradation pattern of Sic1–TAP protein was followed by western blotting, using standard SDS–PAGE. **f**, A wild-type strain (*SIC1wt-TAP*) and a strain also expressing the nondegradable inhibitory domain of Sic1 (*SIC1AN*) under the *GAL1* promoter were arrested in α -factor, followed by the addition of galactose. After 45 min, the cells were washed into galactose media lacking α -factor and Sic1–TAP levels were followed by western blotting (see also Supplementary Fig. 7). **g**, Sic1 mutants with improved Cln–Cdk recognition determinants in suboptimal phosphodegreons trigger rapid degradation of Sic1 in the presence of *SIC1-AN*. Experiment was performed as in panel f.

Cln2–Cdk1. This priming event is followed by docking-enhanced phosphorylations leading to a phosphorylated chain of sites pT5/pT33/pT45/pS76 but no fully phosphorylated paired degrons, as phosphorylation by Cln2–Cdk1 of suboptimal sites in the degrons (T48 and S69, or S80) is slow (Fig. 3a, b). However, the phosphorylated cluster pT5/pT33/pT45/pS76 serves as a powerful Cln2–Cdk1-dependent docking platform for emerging Clb5–Cdk1. As Cln2 levels rise, such priming forces would create a synergistic effect between Cln2 and Clb5, greatly amplifying the impact of low emerging levels of free Clb5–Cdk1 complexes and defining the point of no return for Clb5-dependent positive feedback. A prediction of this model is that changing the limiting suboptimal degron sites to optimal Cdk sites will rescue the lethality of Sic1-1234r1 (Supplementary Fig. 4f), as the degradation in this case should be driven primarily by Cln2. Indeed, changing T48 to a Cdk1 site by introducing a proline at position +1 partially rescued the lethal phenotype of Sic1-1234r1 (Fig. 3d). A similar rescue was attained by introducing a positive determinant for Cln2–Cdk1, an arginine at position +2 (ref. 16), to the site S80. Optimization of the S69 site had no effect. Almost complete rescue was gained by a triple mutation with all three limiting degron sites (T48, S69 and S80) changed to optimal Cdk sites. Importantly, these effects are unlikely to be due to improved binding of phosphodegrons to Cdc4, as the basic residues on the carboxy-terminal side of pS/pT are known to disrupt the Cdc4 interaction¹. Finally, to confirm that degradation of these Sic1 mutants is driven by Cln–Cdk1, instead of Clb5–Cdk1, we additionally mutated the Cln2-specific docking site VLLPP in the triple mutant background (Fig. 3d, lower panel). The *vllpp* mutation abolished the rescue effect of the triple mutant. These data indicate that Cln2 alone does have the potential to drive Sic1 degradation, but the Cln2-driven phosphorylation cascade is terminated at the rate-limiting final steps. However, this mechanism allows creation of the Clb5 docking platform containing the chain of optimal sites pT5/pT33/pT45/pS76.

Finally, to compare further the functions of Clb5 and Cln2, we analysed the degradation of endogenous Sic1. We found that mutation of either the Cln-specific *vllpp* docking motif or Clb5-specific RXL docking sites delayed Sic1 degradation (Fig. 3e), confirming that both Cln2 and Clb5 have a role in the timing of Sic1 degradation. However, when all Clb–Cdk1 activity in the cell was specifically inhibited by overexpression of nondegradable Sic1, endogenous Sic1 was completely stabilized (Fig. 3f and Supplementary Fig. 7), indicating that the key trigger for Sic1 degradation and the G₁/S transition is the emerging free Clb5–Cdk1, after its levels exceed those of the inhibitory complex.

The inability of Cln alone to cause Sic1 degradation could be attributed to the slow phosphorylation rate of sites in the S76/S80 degron. Indeed, by introducing optimal Cln consensus motifs into the slow degron sites, analogously to the experiment described in Fig. 3d, the Sic1 degradation pattern was restored to normal despite the absence of Clb–Cdk1 activity (Fig. 3g). The processive multiphosphorylation cascades, composed of a set of fast and slow steps and different docking specificities, enable this discrimination between the signal outputs of different cyclin–Cdk1 complexes. Furthermore, the Cln output state acts as a primer state for the second complex, creating a potential ‘AND gate’ in which Cln2 is not allowed to trigger the G₁/S transition until sufficient levels of Clb5 activity accumulate.

In conclusion, our model provides novel insights into the multisite phosphorylation mechanism of Sic1 and, potentially, of Cdk1 targets in general. The multiple sites create a network of docking connections that exploit Cks1-dependent and cyclin-specific docking interactions to process Cdk1 signals to achieve proper tuning of the timing of the G₁/S transition (Supplementary Fig. 8). As most Cdk1 targets in the cell contain clusters of multiple sites¹⁸, the regulation of cell cycle switchpoints by Cdk1-dependent multiphosphorylation might prove to be far more complex than generally anticipated, and it is possible

that beneath the seemingly random constellations of phosphorylation sites, an intricate signal processing logic may be hidden.

METHODS SUMMARY

Yeast strains were in the W303 background and are listed in Supplementary Table 3. Plasmid constructs are listed in Supplementary Table 4. Phos-tag Acrylamide AAL-107 (ref. 19) was purchased from NARD Institute.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 February; accepted 20 September 2011.

Published online 12 October 2011.

- Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–521 (2001).
- Borg, M. *et al.* Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Natl Acad. Sci. USA* **104**, 9650–9655 (2007).
- Kim, S. Y. & Ferrell, J. E. Jr. Substrate competition as a source of ultrasensitivity in the inactivation of Wee1. *Cell* **128**, 1133–1145 (2007).
- Thomson, M. & Gunawardena, J. Unlimited multistability in multisite phosphorylation systems. *Nature* **460**, 274–277 (2009).
- Tang, X. *et al.* Suprafacial orientation of the SCF^{Cdc4} dimer accommodates multiple geometries for substrate ubiquitination. *Cell* **129**, 1165–1176 (2007).
- Tyers, M. The cyclin-dependent kinase inhibitor p40^{Sic1} imposes the requirement for Cln G1 cyclin function at Start. *Proc. Natl Acad. Sci. USA* **93**, 7772–7776 (1996).
- Verma, R. *et al.* Phosphorylation of Sic1p by G₁ Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
- Cross, F. R., Schroeder, L. & Bean, J. M. Phosphorylation of the Sic1 inhibitor of B-type cyclins in *Saccharomyces cerevisiae* is not essential but contributes to cell cycle robustness. *Genetics* **176**, 1541–1555 (2007).
- Schneider, B. L., Yang, Q. H. & Futcher, A. B. Linkage of replication to start by the Cdk inhibitor Sic1. *Science* **272**, 560–562 (1996).
- Arvai, A. S., Bourne, Y., Hickey, M. J. & Tainer, J. A. Crystal structure of the human cell cycle protein CksHs1: single domain fold with similarity to kinase N-lobe domain. *J. Mol. Biol.* **249**, 835–842 (1995).
- Hadwiger, J. A., Wittenberg, C., Mendenhall, M. D. & Reed, S. I. The *Saccharomyces cerevisiae* Cks1 gene, a homolog of the *Schizosaccharomyces pombe* *suc1*⁺ gene, encodes a subunit of the Cdc28 protein kinase complex. *Mol. Cell. Biol.* **9**, 2034–2041 (1989).
- Reynard, G. J., Reynolds, W., Verma, R. & Deshaies, R. J. Cks1 is required for G₁ cyclin-cyclin-dependent kinase activity in budding yeast. *Mol. Cell. Biol.* **20**, 5858–5864 (2000).
- Tang, Y. & Reed, S. I. The Cdk-associated protein Cks1 functions both in G₁ and G₂ in *Saccharomyces cerevisiae*. *Genes Dev.* **7**, 822–832 (1993).
- Patra, D., Wang, S. X., Kumagai, A. & Dunphy, W. G. The *Xenopus* Suc1/Cks protein promotes the phosphorylation of G₂/M regulators. *J. Biol. Chem.* **274**, 36839–36842 (1999).
- Deshaies, R. J. & Ferrell, J. E. Jr. Multisite phosphorylation and the countdown to S phase. *Cell* **107**, 819–822 (2001).
- Köivomägi, M. *et al.* Dynamics of Cdk1 substrate specificity during the cell cycle. *Mol. Cell* **42**, 610–623 (2011).
- Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W. & Pavletich, N. P. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell* **26**, 131–143 (2007).
- Holt, L. J. *et al.* Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682–1686 (2009).
- Kinoshita, E., Yamada, A., Takeda, H., Kinoshita-Kikuta, E. & Koike, T. Novel immobilized zinc(II) affinity chromatography for phosphopeptides and phosphorylated proteins. *J. Sep. Sci.* **28**, 155–162 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Kellogg for strains and L. Peil for advice on mass spectrometry. This work was supported by International Senior Research Fellowship No. 079014/Z/06/Z from the Wellcome Trust (M.Lo.), an installation grant from EMBO and HHMI (M.Lo.), no. 1253, grants no. 6766 from the Estonian Science Foundation (M.Lo.) and SF0180071s07 from Estonian Ministry of Education and Research (M.Lo.), EMP grant no. 08071N from the Norwegian government, and grants from the National Institute of General Medical Sciences (D.O.M.) and National Cancer Institute (S.M.R.).

Author Contributions M.K., E.V., R.V., A.I., M.Le. and M.Lo. designed and performed the experiments, except for the isothermal calorimetry experiments, performed by E.R.M.B. and S.M.R.; M.Lo. coordinated the project and wrote the manuscript with assistance from D.O.M. and S.M.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.Lo. (Mart.Loog@ut.ee).

METHODS

Protein purification. TAP-purification of cyclin-Cdk1 complexes was performed as described previously^{20,21} using C-terminally TAP-tagged cyclin constructs cloned into 2 micron vectors and overexpressed from the *GAL1* promoter. For purification of 3HA-Clb2-Cdk1, a yeast strain (a gift from D. Kellogg) with the *GAL1* promoter introduced along with the N-terminal 3HA tag in the chromosomal locus of the *CLN2* gene was used. The overexpressed 3HA-Clb2-Cdk1 complex was purified as described²², using immunoaffinity chromatography with a rabbit polyclonal antibody against the haemagglutinin epitope (purchased from Labas). Purification of N-terminally 6His-tagged recombinant Sic1 constructs was performed using cobalt affinity chromatography. Cks1 was purified as described previously¹².

Phosphorylation assays. For quantitative phosphorylation assays, the substrate concentration was kept in the range of 0.5–2 μM (in the linear $[S]$ versus v_0 range, several-fold below estimated K_M) and the initial velocity conditions were defined as a substrate turnover ranging up to 10%/N of the total concentration of N Cdk sites. The general composition of the assay mixture was as follows: 50 mM HEPES pH 7.4, 5 mM MgCl_2 , 150 mM NaCl, 0.1% NP-40, 20 mM imidazole, 2% glycerol, 2 mM EGTA, 0.2 mg ml⁻¹ BSA, 500 nM Cks1 and 500 μM ATP (with added $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Perkin Elmer)). Around 1–10 nM of the purified kinase complex was used, the amount depending on the setup of the experiment. The optimal working concentration for purified Cks1 was taken as 500 nM according to the optimization performed for cyclin-Cdk1 preparations using Sic1 ΔC as a substrate. For the phosphorylation assay with mutant Cks1, purified kinase complexes were preincubated for 45 min with Cks1 wild type or mutant to compensate for differences in the amounts of Cks1 already present in the preparations. The composition of the preincubation mixture was: 50 mM HEPES pH 7.4, 5 mM MgCl_2 , 150 mM NaCl, 0.2 mg ml⁻¹ BSA and 500 μM ATP. Kinase assay was initiated by adding preincubation mixture and $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ to the substrate. Aliquots were taken at least at two different time points and the reaction was stopped by SDS-PAGE sample buffer. For separating the phosphorylated versions of Sic1, 10% SDS-PAGE was used, supplemented with the Phos-tag reagent¹⁹ according to the instructions from the manufacturer. For quantitative analysis of the dynamics of multisite phosphorylation (in Fig. 1d, e and Supplementary Fig. 2a–d, f) the six major bands separated in Phos-tag SDS-PAGE were quantified as 1- to 6-phosphate species.

For the phosphorylation assay of full-length Sic1 as part of the inhibitory complex, Clb5-TAP-Cdk1 was isolated from yeast cell extract containing overexpressed Clb5-TAP with IgG beads (Supplementary Fig. 4b). The stoichiometric Sic1-Clb5-Cdk1 complex was formed by incubating an excess amount of purified Sic1 with the beads, and unbound Sic1 was removed by washing (50 mM HEPES pH 7.4, 0.5 M NaCl, 0.1% NP-40). The phosphorylation reaction was performed according to the standard kinase assay protocol and was initiated by adding purified cyclin-Cdk1 complexes to the washed beads. The Clb5-dependent phosphorylation of full-length Sic1 as part of the stoichiometric inhibitory complex revealed a similar RXL specificity profile (Supplementary Fig. 4b). This result also suggests that Sic1 ΔC is a valid model substrate, which was chosen for large scale analysis instead of the hard-to-adjust assay with the stoichiometric complex. It is reasonable to consider the N-terminal region of Sic1 as an independent polypeptide entity, given that only a short C-terminal part is required for high-affinity inhibition and that Sic1 is an intrinsically disordered protein²³.

The mutant Cks1 used in our experiments was designed to disrupt the phosphate-binding site and contained the combination of mutations R33E, S82E and R102A (ref. 24). The triple mutants in the hydrophobic patch (hpm) of Clb5 and Clb2 were described previously²⁵, and the hpm of Clb3 (F201A, L205A, T208A) was designed according to sequence homology with other B-type cyclins.

The relative rate constants for different phospho-docking enhanced steps (k_{dock}) were determined using mutated Sic1 ΔC versions containing different pair-wise combinations of Cdk1 sites (the rest of the serine and threonine residues in S/TP motifs were mutated to alanines). It is impossible to estimate the effects of docking on the reaction rates directly, as it is hard to produce a version of substrate protein where the primed docking sites are fully phosphorylated and the secondary sites are unphosphorylated. Therefore, we used an indirect approach by estimating the relative formation rate of doubly phosphorylated species. The k_{dock} was defined as the ratio of the observed fraction of the doubly phosphorylated form and the estimated kinase activity towards the single N-terminal priming site present in the pair-wise Sic1 ΔC construct. The latter parameters were estimated using quantified values for single site specificities from the experiment in Fig. 2b (in the absence or presence of different docking sites). Thus, the k_{dock} values are independent of the rates of the priming steps and reflect only the rates of the secondary steps. The phosphorylation of the substrate was followed in a conventional kinase assay and singly and doubly phosphorylated species were resolved using Phos-tag SDS-PAGE and quantified by PhosphorImager. The k_{dock} was calculated from two

consecutive time points (that is, 8 and 16 min, in the low initial range of total substrate consumption) from at least two independent experiments. We applied a condition of a minimal ratio of 0.7 for the k_{dock} values from these two time points, to ensure that the singly phosphorylated species had not reached the temporary quasi-steady state. The obtained values are 'apparent constants' as it is impossible to precisely determine the relative contributions of processive and distributive mechanisms. Nevertheless, the obtained k_{dock} values provide very good estimates of how much the phosphorylation of a site is enhanced when another site is present.

Western blotting and viability assays. For viability assay, log-phase cultures were equalized in density and spotted as serial dilutions on selective synthetic complete (SC) plates. The plates were incubated for 48–60 h at 30 °C. For the western blotting experiments, cells were grown to OD_{600 nm} = 0.3 and treated for 2.5 h with 1 $\mu\text{g ml}^{-1}$ α -factor or for 2.5 h with 15 mg ml⁻¹ hydroxyurea and released by washing. In the experiments presented in Fig. 3f, g and Supplementary Fig. 7 the *GAL1-SIC1 ΔN* was integrated into the *URA3* locus. In Fig. 2g the Sic1 ΔC -3HA versions were cloned into the pRS315 vector and constitutively expressed under the *ADH* promoter. In the experiments presented in Fig. 3e, f, g the endogenous *SIC1* was C-terminally TAP-tagged. The cells were lysed by bead-beating in lysis buffer containing urea. Blotting of Phos-tag SDS-PAGE gels was performed using a dry system iBlot (Invitrogen). The antibody used for western blotting of 3HA-tagged proteins was HA.11 clone 16B12 from Covance and the antibodies used for western blotting of Cdk1 (Cdc28 (yC-20)) and of TAP-tagged proteins (c-Myc (A-14)) were from Santa Cruz. The rabbit anti-Cks1 antibody was from Labas.

Isothermal calorimetry. Recombinant Sic1 ΔC was phosphorylated with purified Clb2-Cdk1, and complete phosphorylation was confirmed by following the phosphorylation shift by Phos-tag SDS-PAGE. Recombinant Cks1 was expressed in *Escherichia coli* from a pET vector and purified with anion exchange and size-exclusion chromatography. Calorimetry experiments were performed with a VP-ITC system (MicroCal). 0.3–0.4 mM Cks1 was titrated into a 30 μM solution of phospho-Sic1 ΔC . Experiments were carried out at 25 °C in a buffer containing 25 mM Tris and 150 mM NaCl (pH 8.0). Data were analysed with the MicroCal Origin software package. The reported binding constant and stoichiometry are the average from 2 experiments, and the reported errors are the standard deviation of these measurements.

Quantitative mass spectrometry. For quantitative determination of Cks1-dependent phosphorylation of T48, equal amounts of Sic1 ΔC -wt protein were phosphorylated by Clb5-Cdk1 supplemented with normal isotopic ATP ($[\text{O}]\text{ATP}$) or heavy ATP ($[\text{O}]\text{ATP}$) (Cambridge Isotope Laboratories). Kinase assays were incubated at room temperature for 60 min then pooled together in a 1:1 ratio (v/v) in SDS-PAGE sample buffer. The proteins were separated by 10% SDS-PAGE and the gels were stained with Coomassie brilliant blue G-250 (Sigma) and protein bands were excised from the stained gels. Proteins were in-gel digested by LysC/P (10 ng μl^{-1}) (Wako) and peptides were purified by using C18 StageTIPS.

Peptides were separated with an Agilent 1200 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap classic mass-spectrometer (Thermo Electron) equipped with nanoelectrospray ion source (Proxeon). Purified peptides were loaded on a fused silica emitter (75 $\mu\text{m} \times 150 \text{ mm}$) (Proxeon) packed in-house with Reprosil-Pur C18-AQ 3- μm particles (Dr. Maisch HPLC GmbH). Peptides were separated with 30 min 3–40% B gradient (A, 0.5% acetic acid; B, 0.5% acetic acid/80% acetonitrile) at a flow-rate of 200 nl min⁻¹, eluted peptides were sprayed directly into LTQ Orbitrap mass-spectrometer with a spray voltage of 2.2 kV. The mass spectrometry scan range was m/z 300–1,800 and the top 5 precursor ions were selected for subsequent MS/MS scans. A lock-mass was used for the LTQ-Orbitrap to obtain constant mass-accuracy during the gradient analysis. Peptides were identified with the Mascot 2.3 (<http://www.matrixscience.com>) search engine. A peptide mass tolerance of 7 p.p.m. was used and a fragment ion mass tolerance of 0.6 Da. Two missed cleavage sites for LysC/P were allowed. The oxidation of the methionine and the phosphorylation of serine, threonine were set as variable modifications.

- Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
- Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859–864 (2003).
- McCusker, D. *et al.* Cdk1 coordinates cell-surface growth with the cell cycle. *Nature Cell Biol.* **9**, 506–515 (2007).
- Mittag, T. *et al.* Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **18**, 494–506 (2010).
- Bourne, Y. *et al.* Crystal structure and mutational analysis of the *Saccharomyces cerevisiae* cell cycle regulatory protein Cks1: implications for domain swapping, anion binding and protein interactions. *Structure* **8**, 841–850 (2000).
- Loog, M. & Morgan, D. O. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* **434**, 104–108 (2005).

Distinct stem cells contribute to mammary gland development and maintenance

Alexandra Van Keymeulen^{1*}, Ana Sofia Rocha^{1*}, Marielle Ousset¹, Benjamin Beck¹, Gaëlle Bouvencourt¹, Jason Rock², Neha Sharma¹, Sophie Dekoninck¹ & Cédric Blanpain^{1,3}

The mammary epithelium is composed of several cell lineages including luminal, alveolar and myoepithelial cells. Transplantation studies have suggested that the mammary epithelium is maintained by the presence of multipotent mammary stem cells. To define the cellular hierarchy of the mammary gland during physiological conditions, we performed genetic lineage-tracing experiments and clonal analysis of the mouse mammary gland during development, adulthood and pregnancy. We found that in postnatal unperturbed mammary gland, both luminal and myoepithelial lineages contain long-lived unipotent stem cells that display extensive renewing capacities, as demonstrated by their ability to clonally expand during morphogenesis and adult life as well as undergo massive expansion during several cycles of pregnancy. The demonstration that the mammary gland contains different types of long-lived stem cells has profound implications for our understanding of mammary gland physiology and will be instrumental in unravelling the cells at the origin of breast cancers.

The mammary gland is composed of epithelial cells and mesenchymal cells, including adipocytes, fibroblasts, blood vessels and immune cells¹. Initially visible as placode-like structures, mammary glands are specified along the ventral epidermis during embryonic development and progressively invade the underlying mesenchyme, called the mammary fat pad. At puberty, the mammary gland expands considerably to form a highly branched tubular structure that progressively fills the fat pad. During pregnancy, the mammary gland expands further and the terminal end tubular structures differentiate into milk-producing cells. Two main cellular subtypes comprise the mammary gland epithelium: the basal myoepithelial cells and luminal cells, which can differentiate either into ductal cells or milk-producing cells (Supplementary Fig. 1). Whereas alveoli and luminal cells secrete the water and nutrients, the myoepithelial cells, through their contraction, guide the circulation of the milk throughout the ductal tree^{1–3}.

Different assays have been developed to define the differentiation potential of mammary epithelial cells (MECs)^{2,4,5}. *In vitro* assays indicated that both luminal cells and myoepithelial cells can be maintained with their lineage-restricted differentiation potential in a specific medium but only luminal cells can be forced to differentiate into myoepithelial cells upon medium switch⁶. Culture of fluorescence-activated cell sorting (FACS)-isolated human MECs gives rise to either luminal or myoepithelial colonies as well as some bipotent colonies^{6–8}. Culturing MECs as non-adherent cells, called mammospheres, allowed renewal and differentiation of cells with unipotent and bipotent differentiation potential⁹. Transplantation of primary MECs at limiting dilutions suggested the presence of multipotent mammary stem cells and more committed progenitors^{10,11}, and a single MEC is able to reconstitute an entire functional mammary gland in serial transplantation¹⁰. Transplantation of a single FACS-isolated MEC can reconstitute, although at low frequency, a normal mammary gland^{12,13}, indicating that rare multipotent mammary stem cells reside at the top of the cellular hierarchy within the mammary gland. Although

transplantation studies are important to define the differentiation potential of stem cells, these assays mimic a regenerative state that in certain circumstances forces stem cells to differentiate into lineages for which they usually do not contribute to under physiological conditions. For example, hair follicle bulge stem cells give rise to all epidermal lineages upon transplantation and wound healing, but only to hair follicle regeneration under physiological conditions¹⁴. The definitive demonstration that, under physiological conditions, multipotent stem cells are responsible for the development and adult maintenance of the mammary epithelium awaits genetic lineage-tracing experiments³. Here we developed novel lineage-tracing approaches in mice to decipher the cellular hierarchy of the mammary epithelium during physiological conditions.

Multipotent embryonic K14 progenitors

We first assessed the contribution of K14-derived cells to mammary gland development and adult life using K14-Cre/Rosa-YFP mice. The mammary placode arises from the embryonic epidermis at embryonic day 14 (E14). At E17, all MECs expressed K14 and were YFP⁺ in K14-Cre/Rosa-YFP mice and remained YFP⁺ thereafter (Supplementary Figs 2 and 3). During the early stage of mammary gland development, K14 expression encompassed both myoepithelial cells and a fraction of luminal cells (Fig. 1a–c). At birth, the mammary gland consisted of a tubular epithelial structure composed of basal myoepithelial cells expressing K5, K14 and SMA and luminal cells expressing K8 and K19. At the beginning of puberty and thereafter, K14 expression was restricted to the myoepithelial lineage (Supplementary Figs 1 and 2). FACS analysis of mammary gland from K14-Cre/Rosa-YFP mice revealed that all YFP-labelled cells expressed CD24, as previously suggested¹⁵, and could be divided into two populations: CD29^{Hi}CD24⁺ enriched for myoepithelial cells and CD29^{Lo}CD24⁺ enriched for luminal cells^{12,13} (Supplementary Figs 4 and 5). Because K14-Cre is expressed in myoepithelial cells after birth, we investigated whether all MECs derive from embryonic K14⁺ progenitors by administering

¹Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium. ²Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. ³Welbio, Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium.

*These authors contributed equally to this work.

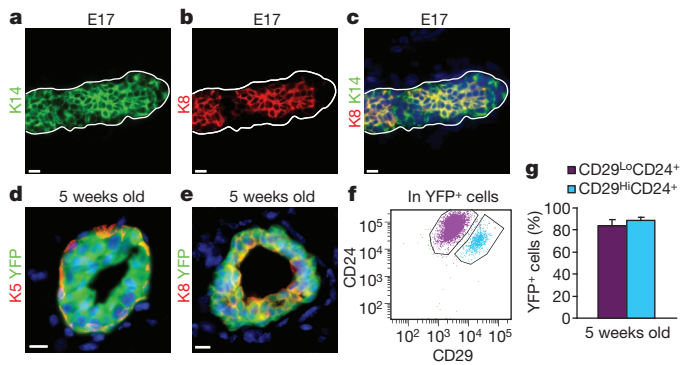


Figure 1 | All mammary epithelial lineages derive from embryonic K14-expressing progenitors. **a–c**, Immunostaining of K8 and K14 in mammary glands at E17. **d, e**, Immunostaining of K5 (**d**) or K8 (**e**) and YFP in K14-rtTA/TetO-Cre/Rosa-YFP MECs 5 weeks after doxycycline administration. **f**, FACS analysis of CD24 and CD29 expression in Lin[−]YFP⁺ cells 5 weeks after doxycycline administration. **g**, YFP expression in Lin[−]CD29^{Lo}CD24⁺ and Lin[−]CD29^{Hi}CD24⁺ populations 5 weeks after doxycycline administration, showing that embryonic K14 tracing marked the vast majority of MECs ($n = 3$ mice). Scale bars, 10 µm; error bars indicate s.e.m.

a single dose of doxycycline to pregnant K14-rtTA/TetO-Cre/Rosa-YFP mice at E17 to label MECs during development (Supplementary Fig. 6). Analysis of the mammary gland at puberty showed that embryonic labelling marked the majority of MECs, including both myoepithelial cells and luminal cells, demonstrating that embryonic K14⁺ progenitors give rise to all MEC lineages (Fig. 1d–g).

K14 myoepithelial stem cells

The transcriptional profiling of CD29^{Hi}CD24⁺ cells revealed that the putative multipotent mammary stem cells^{12,13} are enriched for K5 and K14 (refs 12, 13, 16, 17). To determine whether postnatal K14⁺ cells contain multipotent mammary stem cells, we performed inducible genetic lineage-tracing experiments of K14-expressing cells during puberty and in adult virgin mice. Doxycycline administration over 5 days in K14-rtTA/TetO-Cre/Rosa-YFP mice induced YFP expression in about 40% of myoepithelial cells but did not label luminal cells (Fig. 2a, b, e and Supplementary Figs 7 and 8). Surprisingly, 10 weeks after doxycycline administration, YFP was still exclusively expressed by myoepithelial cells (Fig. 2c–e and Supplementary Figs 7 and 8). To rule out that the remaining CD29^{Hi}CD24⁺ myoepithelial cell population that was not labelled in the previous experiments contains multipotent mammary stem cells, we administered doxycycline continuously to K14-rtTA/TetO-Cre/Rosa-YFP mice during the whole process of pubertal development and found that almost all myoepithelial cells (>97%) but no luminal cells were labelled (Fig. 2f, g and Supplementary Fig. 7m). These data demonstrate that the K14-expressing cells do not contribute to the luminal lineage during mammary gland expansion that occurred during pubertal development.

Administration of a low dose of doxycycline to K14-rtTA/TetO-Cre/Rosa-YFP 4-week-old mice resulted in the labelling of isolated myoepithelial cells 1 week after doxycycline administration. These cells were maintained for several weeks and about 10% of them expanded over time (Fig. 2h, i and Supplementary Fig. 9). YFP⁺ myoepithelial cells expanded further during pregnancy and lactation. Some of them escaped mammary gland involution and reinitiated another cycle of expansion during the following pregnancy and were still present after the second mammary gland involution (Fig. 2j–l and Supplementary Fig. 10). The proportion of YFP-labelled cells was stable over time (Fig. 2l), showing that these unipotent stem cells undergo long-term self-renewal and are not replaced by multipotent stem cells over time.

Doxycycline administration to K14-rtTA/TetO-Cre/Rosa-YFP mice at postnatal day 1 (P1) marked mostly myoepithelial cells and

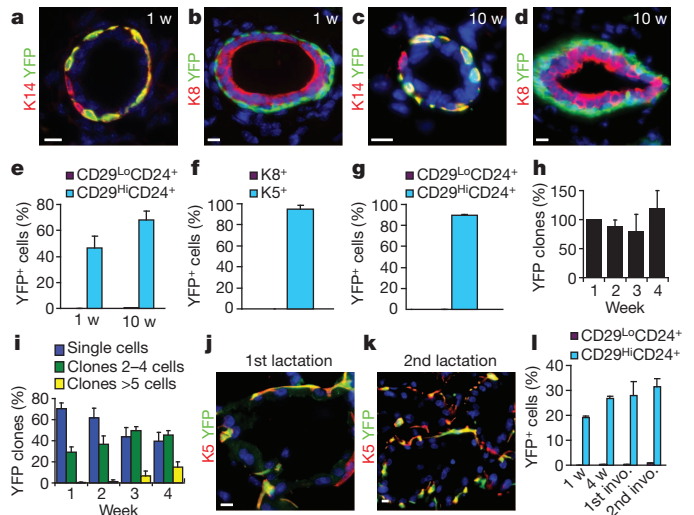


Figure 2 | K14-expressing stem cells ensure mammary myoepithelial lineage expansion and maintenance during puberty, adult life and pregnancy. **a–d**, Immunostaining of K14 (**a, c**) or K8 (**b, d**) and YFP 1 week (**a, b**) and 10 weeks (**c, d**) after doxycycline administration. **e**, Percentage of YFP⁺ cells in Lin[−]CD29^{Hi}CD24⁺ and Lin[−]CD29^{Lo}CD24⁺ populations 1 week and 10 weeks after doxycycline administration at puberty ($n = 3$ mice per time point). **f, g**, YFP expression in myoepithelial cells and luminal cells as determined by immunofluorescence (**f**) and FACS analysis (**g**) after 1 month of continuous doxycycline administration during puberty ($n = 2$ mice per time point). **h**, Percentage of YFP myoepithelial clones observed at different time points. **i**, Temporal distribution of clone size observed as single cell, 2–4 cells and 5 cells or more ($n =$ at least 470 clones from 3 different mice were counted at each time point). **j, k**, Immunostaining of K5 and YFP in MECs during first and second lactation after 1 mg doxycycline administration at puberty. **l**, Percentage of YFP⁺ cells in MECs at different time points after doxycycline administration at puberty ($n =$ at least 2 mice per time point). Scale bars, 10 µm unless stated; error bars indicate s.e.m.

only a small proportion of luminal cells (Supplementary Fig. 11), indicating that the lineage restriction of K14-expressing cells occurred rapidly after birth.

To determine whether K14-rtTA/TetO-Cre targeted a subpopulation of myoepithelial cells distinct from the putative mammary stem cells, we repeated these experiments using K5-CreER mice, another marker preferentially expressed by putative multipotent mammary stem cells^{12,13,16,17}. Tamoxifen administration to P1 and 4-week-old K5-CreER/Rosa-YFP mice initially labelled exclusively myoepithelial cells, giving rise several weeks later to myoepithelial cells only and not luminal cells. Similarly to K14-rtTA/TetO-Cre targeted cells, a fraction of K5-CreER targeted cells was able to clonally expand during puberty and pregnancy (Supplementary Figs 12–15), confirming the presence of myoepithelial stem cells and the apparent absence of multipotent mammary stem cells among postnatal MECs expressing K5 and K14.

Lgr5, a marker of epithelial stem cells in different tissues¹⁸, has been reported to be enriched in the CD29^{Hi}CD24⁺ population¹². Lgr5 is expressed in only 2–3% of MECs, localized to the nipple region. At puberty, the vast majority of Lgr5⁺ cells within the mammary gland were myoepithelial cells, although a small fraction of Lgr5⁺ cells was also luminal (Supplementary Fig. 16a–d). Tamoxifen administration to 4-week-old Lgr5-GFP-CreER/Rosa-Tomato mice preferentially labelled myoepithelial cells, giving rise to myoepithelial cells several weeks later (Supplementary Fig. 16e–m). Functionally, Lgr5⁺ myoepithelial cells were indistinguishable from K14⁺ cells except for their clustering close to the nipple region. The only difference between Lgr5 and K14 is the Lgr5 expression and tracing of rare luminal cells.

K8 luminal stem cells

Lineage-tracing experiments using WAP-Cre mice, which is expressed and active in luminal cells during pregnancy, identified

long-lived cells that display the ability to clonally expand during pregnancy and give rise to luminal and alveolar cells, and are therefore called parity-induced cells¹⁹. *Ex vivo* culture of WAP-Cre mammary explants, in the presence of several growth factors, suggested that cells with similar renewal and differentiation potential as parity-induced mammary progenitors may already exist in nulliparous mice²⁰. To determine whether, under physiological conditions without *ex vivo* manipulations, luminal cells contain stem cells before pregnancy and whether these cells represent unipotent or multipotent stem cells, as has been previously suggested for human MECs⁶, we generated transgenic mice expressing CreER in the luminal lineage using the K8 promoter (Supplementary Fig. 17a). Administration of tamoxifen to K8-CreER/Rosa-YFP mice at P1 only traced luminal cells (Supplementary Fig. 17). Similarly, tamoxifen administration to 4-week-old and adult virgin mice induced YFP expression only in luminal cells (Fig. 3a, b and Supplementary Figs 18 and 19), and after 10 weeks, YFP⁺ cells had expanded but were still luminal cells (Fig. 3c, d and Supplementary Figs 18 and 19), indicating that K8⁺ cells contain luminal stem cells.

Clonal analysis of K8⁺ cells by administration of a low dose of tamoxifen revealed that not all YFP⁺ cells persist long-term and at 4 weeks after tamoxifen administration 40% of YFP luminal clones were lost, indicating that K8-CreER also targets more committed luminal cells (Fig. 3e–h and Supplementary Figs 20–22). Temporal analysis of clone size revealed that about 10% of the YFP clones contained more than 5 YFP⁺ cells, some of which could be even much larger (Fig. 3g and Supplementary Fig. 21), consistent with the targeting of luminal stem cells that clonally expand and participate in luminal cell expansion during puberty and maintenance during adult life.

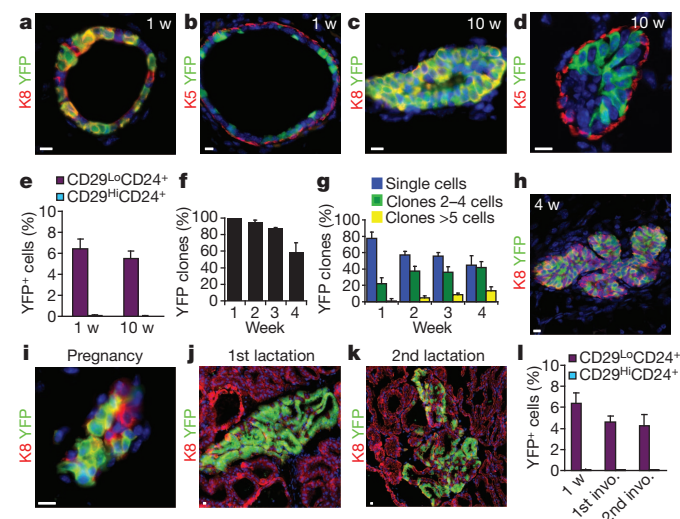


Figure 3 | K8-expressing stem cells ensure mammary luminal lineage expansion and maintenance during puberty, adult life and pregnancy. **a–d**, Immunostaining of K8 (**a**, **c**) or K5 (**b**, **d**) and YFP in MECs 1 week (**a**, **b**) and 10 weeks (**c**, **d**) after 15 mg tamoxifen administration at puberty. **e**, Percentage of YFP⁺ cells in Lin⁺CD29^{hi}CD24⁺ and Lin⁺CD29^{lo}CD24⁺ populations 1 week and 10 weeks after 1 mg tamoxifen administration to K8-CreER/Rosa-YFP mice at puberty ($n = 7$ and 4 mice at 1 week and 10 weeks). **f**, **g**, Number of YFP clones (**f**) and distribution of clone sizes (**g**) observed as single cell, 2–4 cells, and 5 or more cells at different time points after 1 mg tamoxifen administration at puberty ($n =$ at least 526 clones analysed from three different mice at each time point). **h–k**, Immunostaining of K8 (red) and YFP (green) in the mammary gland 4 weeks after induction (**h**), during pregnancy (**i**), lactation (**j**) and second lactation (**k**) in K8-CreER/Rosa-YFP mice treated with 1 mg tamoxifen at puberty. **l**, Percentage of YFP⁺ cells in MECs at different time points after tamoxifen administration in K8-CreER/Rosa-YFP mice treated with 1 mg tamoxifen at puberty ($n =$ at least 2 mice per time point). Scale bars, 10 μ m unless stated; error bars indicate s.e.m.

To establish further the renewal capacities of K8⁺ luminal stem cells and their contribution to milk-producing cells, we induced clonal YFP expression in luminal cells during puberty and followed their fate during pregnancy and lactation (Fig. 3i–l and Supplementary Fig. 23). During pregnancy, only clones of YFP⁺ luminal cells were found. During lactation, very large YFP clones were observed with some lobules that were almost entirely YFP⁺ whereas others were either negative or mosaic for YFP expression (Fig. 3j and Supplementary Fig. 23). K8-CreER-targeted cells differentiated into both luminal and milk-producing cells, as revealed by the co-expression of YFP with K8 and NaPiIb in the fat-milk-producing cells; no myoepithelial cells were YFP⁺ (Fig. 3j and Supplementary Fig. 23d). After involution, some YFP⁺ cells persisted and were able to reinitiate another round of expansion during the following pregnancy and lactation and to escape cell death during the second involution (Fig. 3k and Supplementary Fig. 23). Even after three consecutive cycles of pregnancy and lactation, K8-derived cells were found in luminal cells and milk-producing cells (Supplementary Fig. 23j–m). The percentage of YFP⁺ cells was stable over time (Fig. 3l), indicating that these cells are self-renewing long-term and are not progressively replaced by multipotent stem cells.

K18 committed luminal cells

Different studies suggested that luminal cells are composed of morphologically distinct cell types that are thought to display different proliferation and differentiation capacities^{21–23}. We used another luminal inducible CreER (K18-CreER)²⁴ to determine whether all luminal cells presented similar renewal and differentiation potential as compared to K8-CreER-targeted cells (Supplementary Figs 24–28). Administration of tamoxifen to 4- and 8-week-old K18-CreER/Rosa-YFP mice resulted in a patchy expression of YFP in the mammary gland. Only luminal cells were initially labelled and 10 weeks after tamoxifen administration, YFP-marked luminal cells were still present whereas no myoepithelial cells were YFP⁺ (Supplementary Figs 24 and 25). Clonal analysis during puberty and in adult virgin mice revealed no sign of important clonal expansion, even during pregnancy and lactation (Supplementary Figs 26–28). Together, these data suggest that the luminal cells targeted by K18-CreER display a low cellular turnover, and should be considered as more committed luminal cells, possibly representing the cells that are lost and/or failed to expand in K8 lineage-tracing experiments.

Transplantation and stem cell differentiation

To clarify the discrepancy between the results obtained in transplantation assays^{12,13} and our lineage-tracing experiments, we performed mammary reconstitution assays with cells labelled by our different myoepithelial- and luminal-specific Cre. Transplantation of 10^6 dissociated mammary gland cells—which represent about 5×10^4 living MECs from 4-week-old K14-Cre/Rosa-YFP mice—into the fat pad of NOD/SCID mice allowed the reconstitution of a morphologically normal mammary gland with YFP⁺ myoepithelial cells and luminal cells (Supplementary Fig. 29a–d).

To determine whether the transplantation procedure itself can promote the differentiation potential of YFP-labelled stem cells into the other lineage, we induced YFP expression in myoepithelial cells by administering doxycycline to 4-week-old K14-rtTA/TetO-Cre/Rosa-YFP mice. One week later the mammary gland was dissociated into single cells and a mixture of YFP⁺ myoepithelial cells together with unlabelled luminal cells was transplanted into the mammary fat pad of NOD/SCID mice (Supplementary Fig. 29a). Seven weeks after transplantation, the grafted cells regenerated a new mammary gland expressing YFP in the ducts and growing alveoli (Fig. 4a). Microscopic examination revealed that the vast majority of YFP⁺ cells were myoepithelial cells (Fig. 4b, c and Supplementary Fig. 29e–h) and only very rare clones expressing YFP in both myoepithelial cells and luminal cells were identified (Supplementary Fig. 29i, j and

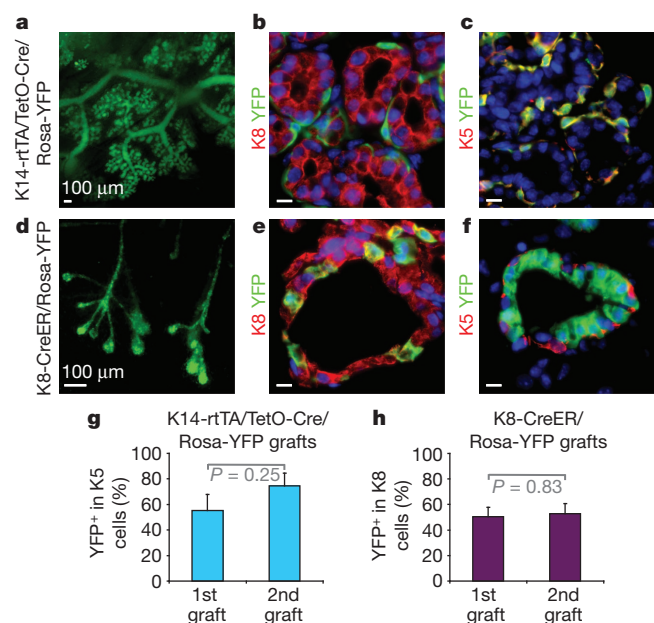


Figure 4 | Myoepithelial and luminal stem cells maintain their unipotent fate in mammary reconstitution assay when co-transplanted in non-limiting conditions. **a**, Whole-mount of the reconstituted mammary gland after transplantation of mammary cells in which myoepithelial cells have been induced to express YFP. **b, c**, Immunostaining of K8 (**b**) or K5 (**c**) and YFP in the K14-rtTA/TetO-Cre/Rosa-YFP graft. **d**, Whole-mount of the reconstituted mammary gland after the transplantation of breast cell suspensions in which a fraction of luminal cells has been induced to express YFP. **e, f**, Immunostaining of K8 (**e**) or K5 (**f**) and YFP in the K8-CreER/Rosa-YFP graft. **g, h**, Percentage of YFP⁺ cells in K5⁺ (**g**) and K8⁺ (**h**) cells in the first and second transplants quantified by immunofluorescence in serial transplantation of unsorted cells from K14-rtTA/TetO-Cre/Rosa-YFP (**g**) and K8-CreER/Rosa-YFP (**h**) mice. Scale bars, 10 μ m unless stated; error bars indicate s.e.m.

Supplementary Table 1). Similar results were obtained after the transplantation of MECs labelled by K5-CreER (Supplementary Fig. 29k–m). Transplantation of YFP-marked luminal cells together with unlabelled myoepithelial cells after tamoxifen administration to K8-CreER/Rosa-YFP mice resulted in the regeneration of a new mammary gland expressing YFP in the duct and growing alveoli, but in this case only luminal cells and not myoepithelial cells were YFP⁺ (Fig. 4d–f). Both luminal and myoepithelial YFP-marked cells could be serially transplanted without a significant decrease in the overall YFP chimaerism and with similar differentiation potential as obtained in the first transplantation. This suggests that both types of stem cells possess a sustained renewal potential and are not progressively replaced by multipotent stem cells (Fig. 4g, h and Supplementary Fig. 30). These experiments clearly demonstrate that luminal and myoepithelial stem cells actively participate in epithelial regeneration in mammary reconstitution assays while still exhibiting the similar lineage-restricted differentiation as observed in unperturbed mammary gland when the two types of unipotent stem cells are present together at non-limiting dilutions.

To determine whether the presence of only one type of mammary stem cell during the transplantation procedure can expand their differentiation potential, FACS-isolated YFP⁺ myoepithelial or luminal cells were transplanted into the mammary fat pad of NOD/SCID mice (Supplementary Fig. 31a). As previously shown^{12,13}, transplantation of FACS-isolated CD29^{Lo}CD24⁺ luminal cells alone failed to reconstitute a new mammary gland upon transplantation whereas CD29^{Hi}CD24⁺ myoepithelial cells alone were able to regenerate a new mammary gland independently of the presence of luminal cells (Fig. 5a–c and Supplementary Table 1). Finally, we assessed whether there is a critical threshold of luminal/myoepithelial cell ratio that

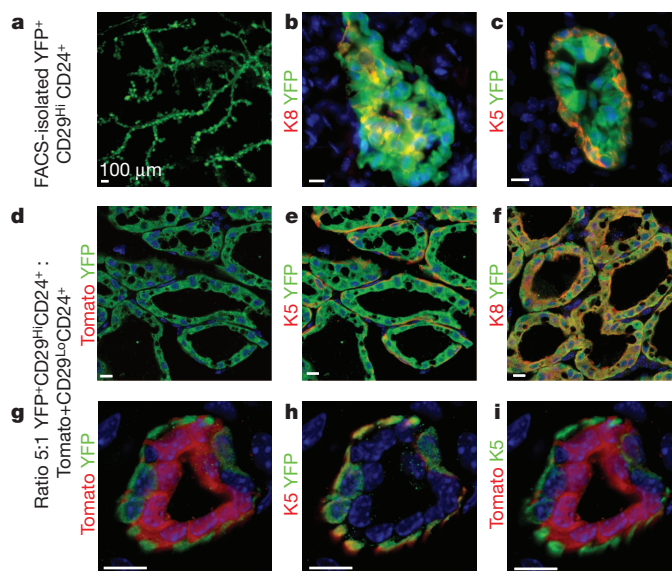


Figure 5 | Myoepithelial but not luminal stem cells can be forced to adopt a multipotent fate in mammary reconstitution assays. **a**, Whole-mount of the graft obtained after the transplantation of FACS-isolated YFP⁺CD29^{Hi}CD24⁺ cells. **b, c**, Immunostaining of K8 (**b**) and K5 (**c**) and YFP. **d–i**, Transplantation of 10⁴ FACS-isolated YFP⁺CD29^{Hi}CD24⁺ cells together with 2,000 Tomato⁺CD29^{Lo}CD24⁺ cells into the mammary fat pads of NOD/SCID mice. Immunostaining of Tomato and YFP (**d, g**), K5 and YFP (**e, h**), K8 and YFP (**f**) and K5 and Tomato (**i**). Scale bars, 10 μ m unless stated.

uncovers the bipotentiality of stem cells. Transplantation of FACS-isolated YFP⁺ myoepithelial cells and Tomato⁺ luminal cells showed that decreasing the luminal/myoepithelial cell ratio to 1/5, which is about ten times lower than physiological conditions, markedly increased the ability of myoepithelial cells to differentiate into luminal cells, as about half of the grafts were exclusively derived from YFP⁺ myoepithelial cells (Fig. 5d–f and Supplementary Table 1), whereas the other half was composed of a mixture of YFP⁺ myoepithelial cells and Tomato⁺ luminal cells (Fig. 5g–i and Supplementary Fig. 31).

Discussion

Our study shows that the mammary gland initially develops from multipotent embryonic K14⁺ progenitors, which give rise to both myoepithelial cells and luminal cells, as has been suggested previously²⁵. During puberty and homeostasis, the expansion and maintenance of each lineage is ensured by the presence of two types of lineage-restricted stem cell, able to differentiate into either myoepithelial or luminal lineages, rather than being maintained by rare multipotent stem cells (Supplementary Fig. 32). Our data cannot rule out that some rare multipotent stem cells that were not targeted by the multiple different Cre lines used in this study exist in the mammary epithelium. However, the long-term maintenance and stable frequency of YFP⁺ cells, their ability to be serially transplanted as well as their massive expansion during pregnancy and lactation clearly show that unipotent stem cells are not replaced by multipotent stem cells over time and that the contribution of multipotent stem cells to mammary gland morphogenesis and adult maintenance, if it exists, is very limited under physiological conditions.

Consistent with our data, previous experiments in which labelled and non-labelled MECs have been transplanted into the fat pad showed that the new mammary glands were composed of mixture of labelled and unlabelled cells^{11,19,26–29}, suggesting that multiple progenitors contribute to mammary gland development and adult maintenance rather than being performed by rare stem cells. In humans, analysis of X-chromosome inactivation indicates that breast epithelium is organized into multiple discrete regions sharing the same inactive X chromosome

and suggested the existence of multiple stem cells scattered throughout the gland³⁰.

Our study also demonstrates that transplantation assays, although extremely informative about the differentiation potential of tissue-specific stem cells, can be misleading in extrapolating the differentiation potential of stem cells under physiological conditions. These results may explain why genetically altered strains of mice that lack basal mammary stem cells in transplantation assays show no particular defects in development or pregnancy^{31,32}. The very low frequency of luminal differentiation of transplanted unipotent myoepithelial stem cells when luminal cells are present within the graft at the physiological ratio, and the increase in the multipotent differentiation of myoepithelial cells when luminal cells were depleted, suggested that either luminal cells restrict the differentiation potential of myoepithelial stem cells or that the differentiation of myoepithelial stem cells into luminal cells is a relatively rare event compared to the natural differentiation of luminal stem cells, and consequently these rare multipotent clones were out-competed by clones originating from luminal stem cells in these non-limited dilution experiments. The reason why myoepithelial cells can adopt a multipotent fate and are able to regenerate a complete mammary gland upon transplantation remains elusive. One possibility would be that mammary reconstitution assays recapitulate the process of mammary gland development and adult K14/K5⁺ stem cells are more prone to dedifferentiate into K14/K5 embryonic multipotent mammary progenitors under these conditions.

METHODS SUMMARY

YFP expression was induced in K8-CreER/Rosa-YFP, K18-CreER/Rosa-YFP and K5-CreER/Rosa-YFP female mice by intraperitoneal tamoxifen injection and in K14-rtTA/TetO-Cre/Rosa-YFP mice by oral administration of doxycycline food or by intraperitoneal injection. Immunostaining was performed as described²⁴. Mammary glands were dissected and lymph nodes removed. Tissues were cut in pieces of 1 mm³ and digested in HBSS plus 300 U ml⁻¹ collagenase plus 300 µg ml⁻¹ hyaluronidase for 2 h at 37 °C under shaking. EDTA was added for 10 min, followed by trypsin-EGTA for 2 min. Cell labelling, flow cytometry and sorting were performed as described^{12,13}. Dead cells were excluded with DAPI; CD45⁺, CD31⁺ and CD140a⁺ cells were excluded (Lin⁻) before analysis. Unsorted mammary cells or FACS-isolated cells were re-suspended in 10 µl PBS and injected into the cleared mammary fat pad of NOD/SCID female mice. Recipient mice were mated 4 weeks after the transplantation and analysed 3 weeks later.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 March; accepted 21 September 2011.

Published online 9 October 2011.

- Watson, C. J. & Khaled, W. T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, 995–1003 (2008).
- Smalley, M. & Ashworth, A. Stem cells and breast cancer: A field in transit. *Nature Rev. Cancer* **3**, 832–844 (2003).
- Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* **23**, 2563–2577 (2009).
- Visvader, J. E. & Smith, G. H. Murine mammary epithelial stem cells: Discovery, function, and current status. *Cold Spring Harb. Perspect. Biol.* **3**, 1–14 (2011).
- Stingl, J. Detection and analysis of mammary gland stem cells. *J. Pathol.* **217**, 229–241 (2009).
- Pechoux, C., Gudjonsson, T., Ronnov-Jessen, L., Bissell, M. J. & Petersen, O. W. Human mammary luminal epithelial cells contain progenitors to myoepithelial cells. *Dev. Biol.* **206**, 88–99 (1999).
- Stingl, J., Eaves, C. J., Kuusk, U. & Emsman, J. T. Phenotypic and functional characterization *in vitro* of a multipotent epithelial cell present in the normal adult human breast. *Differentiation* **63**, 201–213 (1998).
- Gudjonsson, T. *et al.* Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev.* **16**, 693–706 (2002).
- Dontu, G. *et al.* *In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270 (2003).
- Kordon, E. C. & Smith, G. H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921–1930 (1998).
- Smith, G. H. Experimental mammary epithelial morphogenesis in an *in vivo* model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res. Treat.* **39**, 21–31 (1996).

- Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
- Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
- Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature Rev. Mol. Cell Biol.* **10**, 207–217 (2009).
- Sleeman, K. E., Kendrick, H., Ashworth, A., Isacke, C. M. & Smalley, M. J. CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. *Breast Cancer Res.* **8**, R7 (2006).
- Lim, E. *et al.* Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21 (2010).
- Kendrick, H. *et al.* Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* **9**, 591 (2008).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- Wagner, K. U. *et al.* An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* **129**, 1377–1386 (2002).
- Booth, B. W., Boulanger, C. A. & Smith, G. H. Alveolar progenitor cells develop in mouse mammary glands independent of pregnancy and lactation. *J. Cell. Physiol.* **212**, 729–736 (2007).
- Chepko, G. & Smith, G. H. Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue Cell* **29**, 239–253 (1997).
- Rudland, P. S. & Hughes, C. M. Immunocytochemical identification of cell types in human mammary gland: variations in cellular markers are dependent on glandular topography and differentiation. *J. Histochem. Cytochem.* **37**, 1087–1100 (1989).
- Fernandez-Gonzalez, R. *et al.* Mapping mammary gland architecture using multi-scale *in situ* analysis. *Integr. Biol.* **1**, 80–89 (2009).
- Van Keymeulen, A. *et al.* Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis. *J. Cell Biol.* **187**, 91–100 (2009).
- Liu, X. *et al.* Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proc. Natl Acad. Sci. USA* **104**, 12111–12116 (2007).
- Srinivasan, K., Strickland, P., Valdes, A., Shin, G. C. & Hinck, L. Netrin-1/neogenin interaction stabilizes multipotent progenitor cap cells during mammary gland morphogenesis. *Dev. Cell* **4**, 371–382 (2003).
- Naylor, S. *et al.* Retroviral expression of Wnt-1 and Wnt-7b produces different effects in mouse mammary epithelium. *J. Cell Sci.* **113**, 2129–2138 (2000).
- Welm, B. E. *et al.* Sca-1^{pos} cells in the mouse mammary gland represent an enriched progenitor cell population. *Dev. Biol.* **245**, 42–56 (2002).
- Welm, B. E., Dijkgraaf, G. J., Bledau, A. S., Welm, A. L. & Werb, Z. Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. *Cell Stem Cell* **2**, 90–102 (2008).
- Tsai, Y. C. *et al.* Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis. *Cancer Res.* **56**, 402–404 (1996).
- Badders, N. M. *et al.* The Wnt receptor, Lrp5, is expressed by mouse mammary stem cells and is required to maintain the basal lineage. *PLoS ONE* **4**, e6594 (2009).
- Taddei, I. *et al.* β 1 integrin deletion from the basal compartment of the mammary epithelium affects stem cells. *Nature Cell Biol.* **10**, 716–722 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank our colleagues who provided us with reagents, which are cited in the text, and B. Hogan for sharing unpublished mice. We thank our colleagues from the Blanpain laboratory and C. Govaerts for their comments on the manuscript. We thank J. Rosen for discussion and M. Van Lohuizen and K. Nacerddine for their help with the transplantation assay. We thank F. Bollet-Quivogne and J.-M. Vanderwinden for their help with confocal imaging. C.B. and A.V.K. are chercheur qualifié, B.B. is chargé de recherche and M.O. is a collaborateur scientifique of the FRF/FNRS. A.S.R. is supported by TELEVIE and the Portuguese Science Foundation (FCT). N.S. is supported by the Fondation Contre le Cancer. J.R. is supported by the grant F32HL102920. C.B. is an investigator of Wobio. This work was supported by the FNRS, TELEVIE, the program d'excellence CIBLES of the Wallonia Region, a research grant from the Fondation Contre le Cancer, the ULB fondation, the fond Gaston Ithier, a starting grant of the European Research Council (ERC) and the EMBO Young Investigator Program.

Author Contributions C.B., A.V.K., A.S.R. designed the experiments and performed data analysis. A.V.K., A.S.R. and M.O. performed most of the experiments, J.R. generated the K5-CreER knockin mice, B.B., S.D. and A.V.K. performed the FACS analysis and cell sorting. G.B. and N.S. provided technical support. C.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B. (Cedric.Blanpain@ulb.ac.be).

METHODS

Mice. Rosa-YFP³³, Rosa-Tomato³⁴, and Lgr5-GFP-CreER mice¹⁸ were obtained from the Jackson Laboratory. K14-Cre transgenic mice³⁵ and K14-rtTA³⁶ mice were provided by E. Fuchs. TetO-Cre mice³⁷ were provided by A. Nagy. The generation of K18-CreER mice was as previously described²⁴. Mice colonies were maintained in a certified animal facility in accordance with European guidelines. These experiments were approved by the local ethical committee (CEBEA).

Generation of K8-CreER mice. The CreERT2 fragment (supplied by P. Chambon) preceded by the β -globin intron and followed by a SV40 polyA signal was subcloned into pBluescript II SK⁺. The 3.5-kb sequence upstream the ATG codon of the murine K8 gene, obtained from the BAC clone RP23-254K21 (BACPAC Resources Center, Children's Hospital Oakland Research Institute) using the forward primer 5'-GGTGGATCACTTGCCCCCTCCGTTTG-3' and the reverse primer 5'-GGGACAGCGCCAGCGAAGGCC-3', was cloned upstream of the β -globin intron. The resulting K8-CreER fragment of 6.3 kb was released from the backbone by NotI digestion and was microinjected into fertilized oocytes to generate transgenic mice (Jacquemin laboratory). Seven transgenic founders were first identified by PCR, out of 27 mice born. Expression profiles of the K8-CreER founders were screened with reporter Rosa-YFP mice. Four founders expressed the YFP in cells expressing the endogenous K8, and one founder K8 was used throughout this study.

Generation of K5-CreER mice. The CreERT2 fragment, preceded by IRES, was inserted into the 3' UTR of *Krt5* in 129-derived ES cells. Correctly targeted cells were injected into B6 blastocysts. The *neo* selection cassette was removed by crossing a chimaeric K5-CreER male to a β -actin-Flp female (B6;SJL-Tg(ACTFLPe)9205Dym/J). First generation of animals backcrossed to B6 were crossed with Rosa-YFP mice.

Targeting YFP or Tomato expression. K14-Cre/Rosa-YFP female mice express YFP in all cells derived from K14-expressing cells, whereas K14-Cre/Rosa-Tomato mice express Tomato in all cells derived from K14-expressing cells. For lineage tracing induced at 4 weeks or at 8 weeks, K8-CreER/Rosa-YFP, K18-CreER/Rosa-YFP, K5-CreER/Rosa-YFP and Lgr5-GFP-CreER/Rosa-Tomato female mice were induced with 15 mg of tamoxifen (Sigma, diluted in sunflower seed oil, Sigma) by intraperitoneal injection (three injections of 5 mg every other day) and K14-rtTA/TetO-Cre/Rosa-YFP female mice were induced by oral administration of doxycycline food diet (1 g kg⁻¹, BIO-SERV) during 5 days, or 1 month. For lineage tracing induced at E17, pregnant females of K14-rtTA/TetO-Cre/Rosa-YFP mice were fed with a doxycycline food diet at E17 for 3 days. For lineage tracing induced at birth, K8-CreER/Rosa-YFP and K5-CreER/Rosa-YFP newborn mice were induced with 125 μ g tamoxifen (5 μ l of 25 mg ml⁻¹ solution) and K14-rtTA/TetO-Cre/Rosa-YFP female mice were induced with 25 μ g doxycycline (Sigma) (5 μ l of 5 mg ml⁻¹ solution diluted in sterile PBS) by intraperitoneal injection. For two-dimensional clonal analysis, K8-CreER/Rosa-YFP and K18-CreER/Rosa-YFP female mice were induced respectively with 1 mg and 10 mg tamoxifen and K14-rtTA/TetO-Cre/Rosa-YFP female mice were induced with 1 mg of doxycycline by intraperitoneal injection. For clonal analysis by confocal microscopy, K8-CreER/Rosa-YFP and K5-CreER/Rosa-YFP female mice were induced respectively with 0.2 and 7.5 mg tamoxifen by intraperitoneal injection.

Histology and immunostaining. Dissected mammary glands were pre-fixed for 2 h in 4% paraformaldehyde at room temperature. Tissues were washed three times with PBS for 5 min and incubated overnight in 30% sucrose in PBS at 4 °C. Tissues were then embedded in OCT and kept at -80 °C. Sections of 5 μ m (for epifluorescence analysis) or 50 μ m (for confocal analysis) were cut using a HM560 Microm cryostat (Mikron Instruments).

Five-micrometre sections were incubated in blocking buffer (5% NDS/1% BSA/0.2% Triton in PBS) for 1 h at room temperature. The different primary antibody combinations were incubated overnight at 4 °C. Sections were then rinsed three times for 5 min in PBS and incubated with proper secondary antibodies diluted at 1:400 in blocking buffer for 1 h at room temperature. Fifty-micrometre sections were incubated in blocking buffer (5% NDS/1% BSA/0.5% Triton/0.1% Tween 20 in PBS) for 2 h at room temperature. The different primary antibody combinations were incubated overnight at 4 °C. Sections were then rinsed three times in PBS for 45 min and incubated with proper secondary antibodies diluted at 1:400 in blocking buffer overnight at 4 °C. The following primary antibodies were used: anti-GFP (rabbit, 1:1,000, Molecular Probes), anti-GFP (chicken, 1:4,000, abcam), anti-K8 (rat, 1:500, Developmental Studies Hybridoma Bank), anti-K14 (rabbit, 1:2,000, Covance), anti-K14 (chicken, 1:1,000, Covance), anti-K5 (rabbit, 1:1,000, Covance), anti-K19 (rat, 1:500, Developmental Studies Hybridoma Bank), Cy3-coupled-anti-SMA (mouse, 1/1,000, Sigma), NaPiIIb (rabbit, 1/300, provided by J. Biber). The following secondary antibodies were used: anti-rabbit, anti-rat, anti-chicken conjugated to AlexaFluor 488 (Molecular Probes), to rhodamine Red-X or to Cy5 (JacksonImmunoResearch). Nuclei were stained in Hoechst solution and

slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Microscope image acquisition. For whole-mount, freshly dissected mammary glands were squeezed between two coverslips and were imaged using a Zeiss EC Plan-NEOFLUAR $\times 2.5$ objective. Pictures of immunostaining were acquired using $\times 10$ and $\times 40$ Zeiss EC Plan-NEOFLUAR objectives. Pictures were acquired on an Axio Observer Z1 microscope, with the AxioCamMR3 camera and using the Axiovision software (Carl Zeiss). Confocal pictures were acquired at room temperature using a Zeiss LSM780 multiphoton confocal microscope fitted on an Axiovert M200 inverted microscope equipped with C-Apochromat ($\times 40 = 1.2$ numerical aperture) water immersion objectives (Carl Zeiss). Optical sections, $1,024 \times 1,024$ pixels, were collected sequentially for each fluorochrome. The data sets generated were merged and displayed with the ZEN software.

Mammary cell preparation. Mammary glands were dissected and the lymph nodes removed before processing. Tissues were washed in HBSS, and cut in pieces of 1 mm³ with scissors. The tissues were placed in HBSS plus 300 U ml⁻¹ collagenase (Sigma) plus 300 μ g ml⁻¹ hyaluronidase (Sigma) and digested for 2 h at 37 °C under shaking. EDTA at a final concentration of 5 mM was added for 10 min to the resultant organoid suspension, followed by 0.25% Trypsin-EGTA for 2 min before filtration through a 40- μ m mesh, two successive washes in 2% FBS/PBS and labelling.

Cell labelling, flow cytometry and sorting. Two-five million cells per condition were incubated in 250 μ l of 2% FBS/PBS with primary antibodies for 30 min, with shaking every 10 min. Primary antibodies were washed with 2% FBS/PBS, and cells incubated for 30 min with the appropriate secondary antibodies, with shaking every 10 min. Secondary antibodies were washed with 2% FBS/PBS and cells were resuspended in 2.5 μ g ml⁻¹ DAPI (Invitrogen) before analysis. Primary antibodies used were: biotinylated anti-CD24 (1/50, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1/50, clone eBioHmb1-1, eBiosciences), PE-conjugated anti-CD45 (1/50, clone 30-F11, eBiosciences), PE-conjugated anti-CD31 (1/50, clone MEC 13.33, BD Biosciences), PE-conjugated anti-CD140a (1/50, clone APA5, eBiosciences). PE-Cy7 coupled Streptavidin (1/400, BD Biosciences) was used as secondary antibody. Data analysis and cell sorting was performed on a FACSaria sorter using the FACS DiVa software (BD Biosciences). Dead cells were excluded with DAPI; CD45⁺, CD31⁺ and CD140a⁺ cells were excluded (Lin⁻) before analysis of the YFP⁺ cells. For Lgr5-GFP-CreER/Rosa-Tomato and K14-Cre/Rosa-Tomato mice, the same procedure was followed, except that no Lin⁻ exclusion was performed. A minimum of 1,000 YFP⁺ cells was analysed per mouse.

Mammary fat-pad transplantation and analysis. One-million unsorted mammary cells (around 5×10^4 MECs) (for primary grafts), 10^5 unsorted cells from the primary graft (for secondary grafts) or 10,000 sorted cells (either only YFP⁺, or a combination of luminal Tomato⁺ and myoepithelial YFP⁺) as indicated were resuspended in 10 μ l PBS and injected into the number 4 glands of 4-week-old NOD/SCID female mice that had been cleared of endogenous epithelium. Recipient mice were mated 4 weeks after the transplantation, and were killed 3 weeks later, when usually fully pregnant. Recipient glands were dissected, analysed by whole-mount, fixed and placed in OCT for further staining. An outgrowth was defined as an epithelial structure comprising ducts and lobules and/or terminal end buds. The percentage of YFP⁺ luminal or myoepithelial cells was analysed on a minimum of 1,000 YFP⁺ cells on an average of 12 sections per graft. For the grafts of unsorted cells from K8-CreER/Rosa-YFP mice previously treated with 15 mg tamoxifen, a total of 7,885 and 6,591 YFP⁺ cells were analysed for the primary and secondary transplants, respectively. For the grafts of unsorted cells from K14-rtTA/TetO-Cre/Rosa-YFP mice previously treated with doxycycline food for 1 week, a total of 9,352 and of 5,505 YFP⁺ cells were analysed for the primary and secondary transplants, respectively. For the graft of mixed YFP⁺ and Tomato⁺ sorted cells, a total of 4,714 luminal cells was analysed for expression of YFP or Tomato.

Quantification of keratin expression in YFP⁺ cells. A minimum of 20 sections and about 10,000 YFP⁺ cells per mouse were analysed for K5 and K8 expression. For lineage tracing of K14-rtTA/TetO-Cre/Rosa-YFP, a total of 32,853, 31,429, 31,633 and 34,002 cells from three different mice per condition were analysed respectively at 1 week and 10 weeks after tamoxifen administration at 4 weeks, and during first lactation and second lactation. No K8-expressing cells were observed among the YFP⁺ cells in these conditions. For lineage tracing of K8-CreER/Rosa-YFP, a minimum of 10,000 YFP⁺ cells per mouse from three different mice per condition were analysed respectively at 1 week and 10 weeks after tamoxifen administration at puberty, during first lactation and during second lactation. No K5-expressing cells were observed among the YFP⁺ cells in these conditions.

Quantification of the number and size of clones induced in the mammary gland. A minimum of 500 YFP⁺ clones were analysed per time point for each CreER/Rosa-YFP lineage in three independent mice. For each mammary tube section, the number, size (number of cells) and keratin expression (myoepithelial (K5) or luminal (K8)) of YFP clones was scored. The clones were grouped in three

classes according to their size, and their frequency was calculated as the percentage of the total number of clones. The mean number of clones per 1,000 cells was calculated at the different time points studied and normalized as the percentage of the number of clones counted 1 week after tamoxifen administration. For the confocal analysis of size of clones in K8-CreER/Rosa-YFP mice, 69 and 197 clones were analysed respectively 1 week and at 4 weeks after tamoxifen induction.

33. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
34. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
35. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of α -catenin in skin. *Cell* **104**, 605–617 (2001).
36. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
37. Perl, A. K., Wert, S. E., Nagy, A., Lobe, C. G. & Whitsett, J. A. Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl Acad. Sci. USA* **99**, 10482–10487 (2002).